Doctoral theses at NTNU, 2012:63

Jawad Rasool

Analysis and Maximization of Throughput Guarantees Offered in Wireless Networks

NTNU
Norwegian University of
Science and Technology
Thesis for the degree of
Philosophiae Doctor
Faculty of Information Technology, Mathematics
and Electrical Engineering
Department of Electronics and
Telecommunications

NTNU – Trondheim
Norwegian University of
Science and Technology

Doctoral thesis

Jawad Rasool

Doctoral theses at NTNU, 2012:63

NTNU

Jawad Rasool

Analysis and Maximization of Throughput Guarantees Offered in Wireless Networks

Thesis for the degree of Philosophiae Doctor

Trondheim, March 2012

Norwegian University of Science and Technology
Faculty of Information Technology, Mathematics and Electrical Engineering
Department of Electronics and Telecommunications

NTNU – Trondheim
Norwegian University of
Science and Technology

NTNU
Norwegian University of Science and Technology

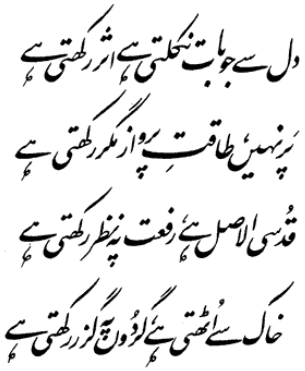
Thesis for the degree of Philosophiae Doctor

Faculty of Information Technology, Mathematics and Electrical Engineering
Department of Electronics and Telecommunications

ISBN 978-82-471-3396-5 (printed ver.)
ISBN 978-82-471-3397-2 (electronic ver.)
ISSN 1503-8181

Doctoral theses at NTNU, 2012:63

Printed by NTNU-trykk



The word springing from the heart surely carries weight,
Though not endowed with wings, it yet can fly in space.
Pure and spiritual in its essence, it pegs its gaze on high,
Rising from the lowly dust, grazes past the skies.

Dr. Muhammad Iqbal
(1877-1938)

Abstract

The field of wireless communications has experienced huge growth in the last decades. Various new user applications have been developed, and the demand for efficient utilization of resources has increased considerably. Opportunistic scheduling schemes can provide higher throughput and increased quality-of-service (QoS) in wireless networks by giving priority to the users with favorable channel conditions.

In this dissertation, one of the main goals is to exploit the channel variations of the user channels to design the scheduling algorithms for improved throughput guarantees. Optimization problems are formulated with an aim at finding optimal scheduling algorithms for maximizing throughput guarantees in a wireless network. We show how the solution to such problems can be obtained both when the throughput guarantees are (i) identical and (ii) different for all the mobile users. We also develop the corresponding adaptive scheduling algorithms, both for the scenarios where single user is scheduled per time-slot and where multiple users are selected in each time-slot (e.g. in MIMO systems). The real-world systems based on Mobile WiMAX, HSDPA, WINNER I and LTE are considered to analyze the proposed scheduling schemes. Results from simulations show that these algorithms can improve the throughput guarantees in modern cellular networks compared to other well-known scheduling algorithms. Another goal is to analyze the performance loss of such scheduling schemes in the case of imperfect channel information at the base station. We also suggest a rate back-off mechanism to reduce the outage probability in that case. Furthermore, we also propose an approximate expression for the throughput guarantee violation probability to analyze the performance of opportunistic scheduling algorithms without conducting experimental investigations. Such an expression can be very useful for the network providers.

Most of the research in joint bandwidth and power allocation for wire-

less multi-user networks has focused on continuous rate, power, and bandwidth allocations in the presence of perfect channel knowledge. However, this is not the case with practical systems. In this dissertation, we therefore also consider the issue of discrete power and bandwidth allocation for discrete-rate multi-user link adaptation with imperfect channel state information. To be more specific, we discuss how the system can be designed in such a scenario for (i) sum rate maximization and (ii) average power minimization in a multi-user setting. The results show that with only a few codes, we can approach the performance of systems that employ continuous (infinitely many) rates. We have also found that the correlation between predicted and actual values of the fading envelope affects the system in the sense that the sum rate is reduced and average power consumption is increased as the correlation is reduced.

Preface

This dissertation is submitted in partial fulfillment of the requirements for the degree of Philosophiae Doctor (PhD) at the Department of Electronics and Telecommunications, Norwegian University of Science and Technology (NTNU). My main supervisor has been Professor Geir E. Øien at the Department of Electronics and Telecommunications at NTNU, while my co-supervisors have been Professor Yuming Jiang at the Department of Telematics at NTNU, and Associate Professor Pål Orten at the University Graduate Center at Kjeller (UniK).

The studies have been carried out at NTNU in the period from October 2007 to October 2011, including one semester of course work, and approximately one year of teaching assistantship duties in different graduate courses.

This work has been funded by the Research Council of Norway (NFR) via the "Spectrum Efficient Communications for future Aeronautical Services" project (ID: 180018/S10), while the teaching assistantship has been funded by the Department of Electronics and Telecommunications, NTNU.

Acknowledgments

I take this opportunity to thank all the people who helped me complete this dissertation.

First and foremost, I would like to express my heartiest gratitude to my supervisor Professor Geir Øien. He was never short of ideas related to my research work. His positive and friendly attitude made it very easy for me to discuss anything with him, both academic and non-academic.

I would like to extend my gratitude to Professors Yuming Jiang and Pål Orten who were my co-supervisors. They were always there to guide me throughout my work. I am also grateful to the project manager Jan Erik Håkegård and Tor Andre Myrvoll from SINTEF, who were always more than helpful to solve the technical and administrative issues. I would specially like to thank Vegard Hassel for his immense cooperation and useful suggestions. I am also indebted to Professors Arvid Næss and Bo Lindqvist from the department of Mathematics at NTNU, for some really useful discussions.

I wish to thank all my friends and colleagues at the Signal Processing Group. I am specially thankful to Mrs. Kirsten Ekseth who is the secretary at our group, and has always been prepared to provide her help. I will miss all the social gatherings that we had in our group.

I am really grateful for the love, prayers and support from my parents and siblings. A lot of thanks to my wife Safia for her love and patience, and for proofreading all my research papers. I would also like to mention my son Shayaan, who makes every day more beautiful than the previous one.

Above all, I thank God Almighty, for giving me the ability and strength to complete this research successfully.

Jawad Rasool
December 2011, Trondheim

Contents

List of Figures

List of Tables

Abbreviations

ACM	Adaptive Coding and Modulation
AMTGS	Adaptive Maximum Throughput Guarantee Scheduling
ASSE	Average System Spectral Efficiency
ATC	Air Traffic Control
AWGN	Additive White Gaussian Noise
BER	Bit Error Rate
BC	Broadcast Channel
CAC	Call Admission Control
CDF	Cumulative Distribution Function
CDM	Code-Division Multiplexing
CDMA	Code-Division Multiple Access
CNR	Carrier-to-Noise Ratio
CSI	Channel State Information
CSMA/CA	Carrier-Sense Multiple Access with Collision Avoidance
DPC	Dirty Paper Coding
DRRM	Dynamic Radio Resource Management
FDD	Frequency-Division Duplexing
FDM	Frequency-Division Multiplexing

FDMA	Frequency-Division Multiple Access
FIFO	First In First Out
HOL	Head-Of-the-Line
HSDPA	High Speed Downlink Packet Access
HSPA	High Speed Packet Access
JFI	Jain's Fairness Index
LDT	Large Deviation Theory
LoS	Line-of-Sight
LTE	Long Term Evolution
LWDF	Largest Weighted Delay First
MAC	Medium Access Control
MAP	Maximum A Posteriori
MCS	Maximum Carrier-to-noise ratio Scheduling
MIMO	Multiple-Input, Multiple-Output
ML	Maximum Likelihood
MLWDF	Modified Largest Weighted Delay First
MTGS	Maximum Throughput Guarantee Scheduling
MUD	Multi-User Diversity
NCS	Normalized Carrier-to-noise ratio Scheduling
NORR	Normalized Opportunistic Round Robin
OFDM	Orthogonal Frequency-Division Multiplexing
OFDMA	Orthogonal Frequency-Division Multiple Access
ORR	Opportunistic Round Robin
PDA	Personal Digital Assistant

PDF	Probability Density Function
PFS	Proportional Fair Scheduling
QoS	Quality-of-Service
RR	Round Robin
RRM	Radio Resource Management
SDM	Space-Division multiplexing
SDMA	Space-Division Multiple Access
SINR	Signal-to-Interference-plus-Noise Ratio
SRRM	Static Radio Resource Management
TDD	Time-Division Duplexing
TDM	Time-Division Multiplexing
TDMA	Time-Division Multiple Access
TG	Throughput Guarantee
TGVP	Throughput Guarantee Violation Probability
ULA	Uniform Linear Array
VoIP	Voice over IP
WiMAX	Worldwide Interoperability for Microwave Access
WINNER	Wireless world Initiative New Radio

Notations

$\lvert\cdot\rvert$	Absolute value
$\lVert\cdot\rVert$	Norm
$(\cdot)^{-1}$	Matrix inversion
$\mathbf{x}^H$	Hermitian conjugate
$\mathbf{x}^{\mathbb{T}}$	Transpose
α_i, β_i	Constants to be optimized
β_c	Minimum CNR within interval c in an ACM scheme
γ	CNR
$\gamma_b, \gamma_b(\epsilon)$	Back-off CNR
$\gamma_{c,l}$	CNR thresholds
$\gamma_i, \gamma_i(t), \gamma_i(n)$	CNR of user i
$\gamma_{i,j}$	CNR of user i in jth time-slot he is scheduled
$\bar{\gamma}_i$	Average CNR of user i
$\bar{\gamma}_t$	Total average CNR (averaged over all the users)
$\hat{\gamma}_{c,l}$	Switching thresholds
$\hat{\gamma}_i, \hat{\gamma}_i(n), \hat{\gamma}_i(t)$	Predicted CNR of user i
Δ	Prediction delay
η_i	SINR of user i

$\eta_{i,m}$	SINR of user i on beam m
$\vartheta_i, \vartheta_i(n)$	Adaptive parameter in AMTGS1 algorithm
κ_l	lth order cumulant
$\lambda(\cdot)$	Cumulant generating function
$\mu_{b_{i,j}}$	Mean of $b_{i,j}$
ρ_i	Normalized correlation between the actual and predicted CNRs
$\sigma^2_{b_{i,j}}$	Variance of $b_{i,j}$
τ_{max}	Maximum delay spread
$v_i, v_i(n)$	Adaptive parameter in AMTGS2 algorithm
ϕ_i	Priority given to user i
$\Psi(x)$	Refer to (2.21)
ω_i	Fraction of Bandwidth allocated to user i in an ACM
Ω_i	Expectation of fading gain of user i
a_p	Amplitude of the pilot-symbols
b	Actual number of bits provided by the system
b_i	Actual number of bits of user i
$b_{i,j}$	Bits transmitted to/from user i in the jth time-slot
$\bar{b}_{i,k}$	average number of bits of user i in k time-slots
B	Promised throughput guarantee in bits
B_c	Coherence bandwidth
B_i	Promised throughput guarantee of user i in bits
B_{in}	Total number of bits assigned to user i after n time-slots
B_s	Signal bandwidth

c	Speed of light
$\mathbf{c}_\Delta$	Vector of filter coefficients
$\mathbf{c}_{\Delta,\mathrm{MAP}}$	MAP optimal filter coefficients
C	Number of rates (codes) in an ACM scheme
C_e	Euler's constant
$E_1(\cdot)$	Exponential integral function of first order
$\mathbb{E}[\cdot]$	Expectation operator
$\mathbb{E}_K[\cdot]$	Expectation over a time-window containing K time-slots
f_c	Carrier frequency
f_D	Maximum Doppler frequency shift
$f_X(x)$	PDF of the random variable X
$f_X(x\|i)$	PDF of the random variable X of the selected user i
$F_X(x)$	CDF of the random variable X
$F_X(x\|i)$	CDF of the random variable X of the selected user i
$g_i(t), g_i(n)$	Channel fading gain of user i
$h_i(t), h_i(n)$	Complex-valued channel response of user i
$\tilde{h}_i(n)$	ML estimate of $h_i(n)$
$\hat{h}_i(n)$	Predicted channel of user i
$\tilde{\mathbf{h}}_i(n)$	Vector of ML estimates
i	User index
$i^*, i^*(n)$	Index of the selected user
$i_m^*, i_m^*(n)$	Index of the selected user on beam m
$I(\cdot)$	Large deviation rate function
$I_0(\cdot)$	Zeroth-order modified Bessel function of the first kind

$\mathbf{I}_{\mathcal{K}\times\mathcal{K}}$	A $\mathcal{K} \times \mathcal{K}$ identity matrix
$J_0(\cdot)$	Zeroth-order Bessel function of the first kind
k	Time-slots allocated to a user
K	Total number of time-slots in the time-window T_W
$\mathcal{K}$	Filter length
ln	Natural logarithm
L	Number of power levels per code in an ACM scheme
$\mathcal{L}$	Spacing between adjacent pilot-symbols
m_l	lth order moment
M	Number of antennas at the base station
$M(\cdot)$	Moment generating function
$\max(\cdot)$	Maximum operator
mod	Modulo operator
$\min(\cdot)$	Minimum operator
n	Discrete time index
$n_i(t), n_i(n)$	Complex-valued AWGN noise of user i
N	Number of users in the network
N_0	Noise power spectral density
N_p	Number of propagation paths
$p(i)$	Access probability of user i
$p_i(n)$	Deterministic pilot-symbols for user i
$\tilde{p}_i(n)$	Received pilot-symbols at user i
$p_K(k\|i)$	Probability that user i gets k out of K time-slots
$p_{out}(i)$	Outage probability of user i

P	Transmit power
$\bar{P}$	Average transmit power
$\mathcal{P}_{c,i}$	Probability of selecting code c for user i
$\Pr[\cdot]$	Probability
$q_b(a,\epsilon)$	Inverse of the complementary Marcum-Q function with respect to its second argument
$Q(a,b)$	Marcum-Q function
r	Rate
$r_i, r_i(n)$	Rate of user i
$r_{i,m}, r_{i,m}(n)$	Rate of user i on beam m
$\hat{r}_i, \hat{r}_i(n)$	Predicted rate of user i
$\mathbf{r}_{\Delta,\mathcal{K}}$	Vector of $\mathcal{K}$ elements representing the correlation between the fading to be predicted and the fading at pilot-symbol instants
$\bar{R}_i$	Average rate of user i
$R(\tau)$	Autocorrelation function
$\mathbf{R}_{\mathcal{K}}$	Autocorrelation matrix of the fading at pilot-symbol instants
$\mathcal{S}$	Set of selected users
SE_c	Spectral efficiency of code c
t	Continuous time variable
t_c	Length of exponentially weighted window in PFS algorithm
T_c	Coherence time
T_i	Accumulated time of user i over the time-window T_W
$\mathcal{T}_i$	Average throughput of user i in time-window t_c
T_s	Symbol duration

T_{TS}	Length of time-slot
T_W	Length of time-window
TGVP_i	TGVP of user i
v	Speed of the user
$\mathbf{w}_m$	$M \times 1$ beamforming vector
$\mathbf{w}_{i^*}$	Beamforming vector of the selected user
W	Transmission bandwidth
$W_i(n)$	HOL packet delay of user i
$\mathbf{W}$	Beamforming matrix
$x_i(t), x_i(n)$	Complex-valued transmitted signal of user i
$\mathbf{x}$	$M \times 1$ transmitted signal
$y_i, y_i(t), y_i(n)$	Complex-valued received signal of user i

Chapter 1

Introduction

1.1 Motivation

The field of wireless communications has experienced huge growth in the last decades. Various user applications have been developed, for example voice over IP (VoIP), video conferencing and online-banking. These applications have widely contrasting requirements in terms of throughput, delay, and reliability. Furthermore, the demand for efficient utilization of radio resources has also increased considerably. The resources in question can be, for example, bandwidth, time-slots, frequencies, or code-sequences. *Scheduling* is a process of selecting the mobile users that are going to transmit or receive information on the wireless channel in order to efficiently utilize the radio resources, and fulfill certain quality-of-service (QoS) requirements [Fattah and Leung, 2002]. Practically, it is hard to fulfill all the requirements when designing a scheduling algorithm. In modern wireless networks, *opportunistic multi-user scheduling* has been implemented in order to obtain a more efficient utilization of the scarcely available radio spectrum. Opportunistic scheduling schemes can provide higher throughput and increased QoS in wireless networks by giving priority to users with favorable channel conditions. For wireless networks carrying real-time traffic, it is in the interest of both the customer and the network operator to have more exact QoS measures. One such measure is a *throughput guarantee*, which will be defined formally in Section 1.8.1. Essentially, it is a guarantee that a user will transmit/receive a certain amount of data within the time-window. With regard to throughput guarantees, our motivation

in this thesis is to find the answer to the following questions:

- How can we quantify the throughput guarantees for a certain scheduling algorithm without conducting experimental investigations?
- How can we maximize the throughput guarantees offered in wireless networks with
  - identical throughput guarantees for all the users, and/or
  - different throughput guarantees for all the users,

  when
  - a single user is scheduled per time-slot?
  - multiple users are selected in a single time-slot?
- What is the impact of imperfect channel information at the base station on the offered throughput guarantees?
- How can the outage probability be controlled in the case of imperfect channel information at the base station?

In practical wireless networks, the available transmission power and bandwidth are limited resources. Therefore, joint bandwidth and power allocation for wireless multi-user networks is essential in order to improve the network performance. Most of the research in joint bandwidth and power allocation for wireless multi-user networks has focused on continuous rate, power, and bandwidth allocations in the presence of perfect channel knowledge. However, this is not the case with practical systems. In this dissertation, we therefore would also like to find the answers to the following questions:

- How can the system with discrete power and bandwidth allocation for discrete-rate multi-user link adaptation be designed for
  - sum rate maximization, and
  - average power minimization?
- What is the impact of imperfect channel prediction on the performance of such a system?

The remainder of the chapter is organized as follows. In Section 1.2 and Section 1.3, we briefly discuss the characteristics of cellular wireless network and wireless channel respectively. Main components of radio resource management are described in Section 1.4. Various scheduling algorithms and the criteria to assess their performance are discussed in Sections 1.5-1.8. A brief literature survey is given in Section 1.9. In Section 1.10, an outline of the rest of the dissertation is provided. A short summary of the papers not included in this thesis is given in Section 1.11.

1.2 Cellular/Centralized Wireless Networks

A Cellular network architecture is typically divided into small sections called *cells*. Within each cell, there is a fixed location transceiver known as a *base station* with limited range. A cellular network also consists of mobile stations, which are served by the base station. The mobile station is a device used by the user, which can both transmit and receive information – this is usually a mobile phone, but it can also be, for example, a personal digital assistant (PDA), a computer, or a wireless sensor. Such an architecture is often termed as *centralized* since most of the network activity is managed by the network infrastructure, i.e., base stations, base station controllers, and switches. An illustration is given in Figure 1.1.

In such a centralized system, the transmission from the base station to the mobile users is denoted as *downlink* transmission and the transmission from the mobile users to the base station is termed as *uplink* transmission. Most communication systems are bi-directional. Duplexing techniques are used to separate the uplink and downlink transmissions from each other. In *time-division duplexing* (TDD), the time-slots are used to separate uplink and downlink transmission. Another possibility is *frequency-division duplexing* (FDD), which assigns separate frequencies for uplink and downlink transmission [Goldsmith, 2005].

1.3 Wireless Channel

1.3.1 Main Features

The strength of a wireless signal varies both in time and frequency. This variation of the signal power can be attributed to the *path loss* caused by

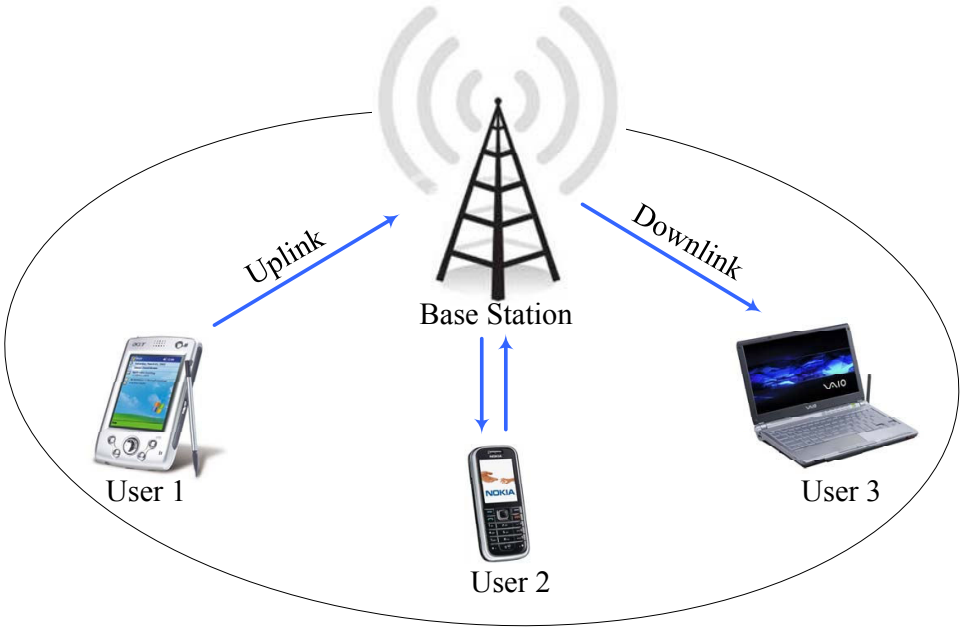


FIGURE 1.1: Centralized Wireless Communication

the attenuation due to distance between the base station and the mobile user, *shadowing* caused by signal energy loss due to obstacles e.g. buildings, trees etc., and *multipath fading* which results in rapid fluctuations of the signal, and is caused by constructive and destructive interference of multiple signal paths between the base station and the mobile user.

The fading is said to be *slow* if many successive symbols are affected by a particular fade level, i.e. the symbol duration T_s [seconds] is much smaller than the coherence time T_c [seconds] of the channel; otherwise it is *fast* fading.

If all the spectral components of the signal are affected in a similar manner, the fading is termed as *flat*. In this case, the signal bandwidth B_s [Hz] is much smaller than the coherence bandwidth B_c [Hz] of the channel. On the other hand, if the spectral components of the signal are affected differently, the fading is *frequency selective* [Rappaport, 2002].

1.3.2 Fading Channel Model

The multipath fading channel model between the base station and user i is shown in Figure 1.2.

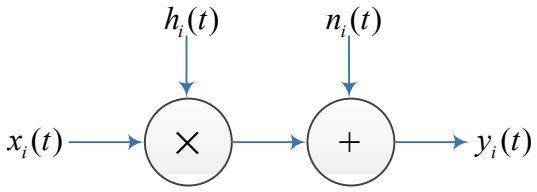


FIGURE 1.2: Channel Model

The received signal $y_i(t)$ can be written as

$$y_i(t) = h_i(t) \cdot x_i(t) + n_i(t), \tag{1.1}$$

where $x_i(t)$ is the complex-valued transmitted signal, $n_i(t)$ is complex-valued additive white Gaussian noise (AWGN), and $h_i(t)$ is the complex-valued channel response. Due to fading, the received carrier amplitude is modulated by the fading gain $g_i(t) = |h_i(t)|$, with expectation $\Omega_i = \mathbb{E}[g_i^2(t)]$. We define the instantaneous carrier-to-noise ratio (CNR) as [Simon and Alouini, 2005]

$$\gamma_i(t) = \frac{g_i^2(t)P}{N_0 W}, \tag{1.2}$$

and the average CNR as

$$\bar{\gamma}_i = \frac{\Omega_i P}{N_0 W}, \tag{1.3}$$

where P [Watt] is the transmit power, N_0 [Watt/Hz] is the one-sided power spectral density of AWGN noise, and W [Hz] is the transmission bandwidth.

The *Rayleigh fading model* is commonly used to model multipath fading with no direct line-of-sight (LoS) path. In this case, the probability density function (PDF) of the CNR γ_i is given by [Simon and Alouini, 2005]

$$f_{\gamma_i}(\gamma) = \frac{1}{\bar{\gamma}_i} e^{-\frac{\gamma}{\bar{\gamma}_i}}. \tag{1.4}$$

The temporal correlation of the channel is both dependent on the speed v [m/s] of the user and on the carrier frequency f_c [Hz] of the channel. For most of the simulations in this dissertation, we assume (modified) Jakes' correlation model defined in [Zheng and Xiao, 2003]. The flat Rayleigh fading channel response $h_i(t)$ can in this case be modeled as sum-of-sinusoids, and is given as

$$h_i(t) = h_I(t) + jh_Q(t), \tag{1.5}$$

with

$$h_I(t) = \frac{2}{\sqrt{N_p}} \sum_{m=1}^{N_p} \cos(\varphi_m) \cos\left(2\pi f_D t \cos\left(\frac{\pi(2m-1)+\theta}{4N_p}\right) + \phi\right),$$

$$h_Q(t) = \frac{2}{\sqrt{N_p}} \sum_{m=1}^{N_p} \sin(\varphi_m) \cos\left(2\pi f_D t \cos\left(\frac{\pi(2m-1)+\theta}{4N_p}\right) + \phi\right),$$

where

$$f_D = \frac{v f_c}{c} \text{ Hz} \tag{1.6}$$

is the maximum Doppler frequency shift, c [m/s] is the speed of light, φ_m, ϕ and θ are uniformly distributed over $[-\pi, \pi]$ for all m, and N_p is the number of propagation paths. In Figure 1.3, typical realizations of the channel gain $g_i^2(t) = |h_i(t)|^2$ are plotted for different values of f_D. We see that the channel gain changes much slowly for smaller f_D.

1.4 Radio Resource Management

Many different wireless standards have been developed during the last decades [Freescale Semiconductor, 2007; Nomor Research: White Paper, 2006; Sternad, Svenson, and Klang, 2006; WiMAX Forum, 2006]. The characteristics of these standards depend on the objectives of the wireless network design. For example, the main design objectives for cellular networks can be high throughput and fulfillment of the QoS requirements of the mobile users. To be able to provide QoS to the users in a wireless network, *radio resource management* (RRM) should be implemented. RRM can be defined as the process of controlling the radio resources with an aim to maximize the overall system performance [Hasu, 2007; Zander, 1997; Zhang, Hu, and Fujise, 2007].

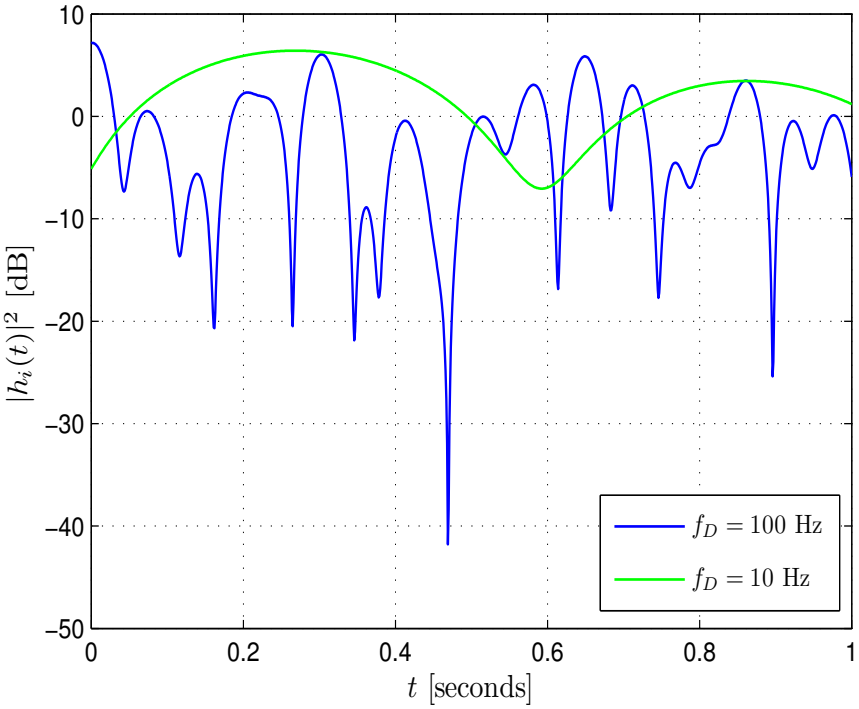


FIGURE 1.3: Channel gain using modified Jake's model

1.4.1 Static Radio Resource Management

Static radio resource management (SRRM) deals with the management and planning of fixed system characteristics. Three important fixed characteristics are the system architecture, the available frequency spectrum, and the access techniques implemented in the system.

1.4.1.1 System Architecture

The main issues to be considered while planning the system architecture are the location of the base stations, the number and type of antennas per base station or mobile terminal, the complexity of the equipment, and the available transmission power.

1.4.1.2 Frequency Spectrum

In most countries the frequency spectrum is publicly regulated to limit the interference between different systems. This means that each network operator governs his own share of the frequency spectrum and can deploy wireless networks that operate on these frequencies. There are also unlicensed parts of the frequency spectrum that can be used freely.

1.4.1.3 Multi-user Access Schemes

Access schemes are implemented so that multiple users can share the same physical channel, and they are designed to exploit the architecture and the available frequency spectrum in the most efficient way. Commonly used access techniques are ALOHA-based techniques (e.g. CSMA/CA), time-division multiple access (TDMA), frequency-division multiple access (FDMA), code-division multiple access (CDMA), or space-division multiple access (SDMA). In TDMA, the users share the resource over time and only one user transmits or receives at a particular time. In FDMA, the users transmit and receive simultaneously, but over different frequency bands. In CDMA the users are not separated in either time or frequency, but by the use of codes. SDMA uses direction (using directional antennas) which can be channelized and assigned to different users [Goldsmith, 2005].

1.4.2 Dynamic Radio Resource Management

Due to the temporally and spatially varying channel quality of each of the mobile users in the network, the efficiency of the network will also vary with time. In addition, a wireless network in full use will also have a constantly changing number of mobile users, each having changing number of applications needing to transmit or receive data over the network. *Dynamic radio resource management* (DRRM) is related to the management of these constantly changing characteristics of the wireless network.

In the following, we discuss some of the parts of the DRRM briefly.

1.4.2.1 Link Adaptation

Link adaptation, or adaptive coding and modulation (ACM), refers to a set of techniques where the modulation, coding rate, transmit power and/or other signal parameters are adapted according to the changing channel

conditions. For example, we can adapt the transmit power to have a constant received CNR, or we can transmit at higher data rates when the channel conditions are more favorable [Goldsmith, 2005].

The base station requires the channel state information (CSI) from all the mobile users for the process of link adaptation. This information can, for example, be obtained from the deterministically known *pilot-symbols* transmitted in-between the data symbols [Duong and Øien, 2007].

A wide range of link adaptation techniques have been developed during the last decades and most modern wireless systems and standards have implemented such techniques [Alouini and Goldsmith, 2000; Cavers, 1972; Gjendemsjø, Øien, Holm, Alouini, Gesbert, Hole, and Orten, 2008; Goldsmith and Chua, 1997; Hole, Holm, and Øien, 2000; Holm, Øien, Alouini, Gesbert, and Hole, 2003; Svensson, Øien, Alouini, and Sampei, 2007; Tao and Czylwik, 2011; Webb and Steele, 1995].

1.4.2.2 Call Admission Control

Fulfilling the QoS for all the users in wireless networks is a challenging problem due to the scarcity of wireless resources and mobility of the users. Call admission control (CAC) is a mechanism that restricts the access to the network based on resource availability in order to prevent network congestion and service degradation for already selected users [Fang and Zhang, 2002; Ghaderi and Boutaba, 2006]. A new user is not admitted if there are not enough resources to meet the QoS requirements of the new user without violating the QoS for already accepted users. This denial of admission is termed as *call blocking* [Ghaderi and Boutaba, 2006]. If a user's application is degrading the QoS of many other users, this troublesome user should perhaps be dropped from the network. This procedure is often called *call dropping*. In general, call dropping is considered to have a more negative impact from the user's point of view than call blocking. A good CAC scheme should balance the call blocking and the call dropping in order to provide the desired QoS requirements [Fang and Zhang, 2002].

1.4.2.3 Mobility Management

The goal of the mobility management is to support mobile users enjoying their calls while simultaneously roaming freely without the disruption of communications.

There are two main aspects of the mobility management; location management and handoff management [Sun, Howie, and Sauvola, 2001]. *Location management* concerns how to locate a mobile user, track its movement, and update the location information. When a mobile user moves from one cell to another, the call may have to be handed off to another cell. *Handoff management* focuses mostly on the issues related to this handoff.

1.4.2.4 Scheduling

In a cellular/centralized network (Figure 1.1), the scheduling decision is taken by the base station both for the downlink and uplink transmission. The downlink transmission is carried out by the base station while each mobile user performs the uplink transmission. Therefore, the base station needs to distribute the scheduling decision to the scheduled users for each time-slot in the uplink.

There are two main categories of scheduling, opportunistic and non-opportunistic, which will be discussed in the Sections 1.5 and 1.6 respectively.

1.5 Non-Opportunistic Scheduling

Non-opportunistic scheduling algorithms do not take channel conditions into account when selecting a user to transmit or receive in the next time-slot. A few examples of such algorithms are round robin scheduling, first in first out scheduling, and largest weighted delay first scheduling.

1.5.1 Round Robin Scheduling

One of the simplest schedulers used in TDMA systems is *round robin* (RR) scheduler which schedules the users in a sequence, one after the other. Thus each user is assigned equal number of time-slots in general. The RR scheduling algorithm for selecting a single user in each time-slot can be formulated as follows:

$$i^*(n) = n \bmod N, \tag{1.7}$$

where $i^*(n)$ denotes the index of the selected user at time-slot n, N is the number of users in the network, and mod is the modulo operator which returns the remainder of the division of one number by another.

Round robin scheduling is easy to implement and starvation-free. The disadvantage with this type of scheduling is that if a user has no data to send, it still gets the channel for fixed number of time-slots and wastes system resources.

1.5.2 First In First Out Scheduling

First in first out (FIFO) scheduler is another simple scheduler that schedules the users according to their waiting times. The user i whose head-of-the-line (HOL) packet has spent the largest time at the base station is selected for transmission in time-slot n:

$$i^*(n) = \underset{1 \le i \le N}{\operatorname{argmax}} W_i(n), \tag{1.8}$$

where $W_i(n)$ [seconds] is the HOL packet delay in user i's buffer at the beginning of time-slot n. This algorithm can be used both for uplink and downlink scheduling since W_i can denote the delay of the HOL packets in either the users' output buffers on the uplink or the buffers at the base station containing packets for downlink transmission to each of the mobile users.

1.5.3 Largest Weighted Delay First Scheduling

In [Stolyar and Ramanan, 2001], *largest weighed delay first* (LWDF) scheduling algorithm is proposed which selects a user whose weighted delay $\phi_i W_i(n)$ is maximum.

$$i^*(n) = \underset{1 \le i \le N}{\operatorname{argmax}} \phi_i W_i(n), \tag{1.9}$$

where $W_i(n)$ is defined as above, and ϕ_i is a constant denoting the priority given to user i. If all the users are given equal priority, LWDF scheduler reduces to FIFO scheduler.

1.6 Opportunistic Scheduling

Opportunistic Scheduling algorithms take channel conditions into account when selecting a user. The mobile user that has the best channel conditions, as measured by some metric, is given priority to transmit or receive data in the next time-slot. By giving priority to this user, the system performance (based on a certain QoS measure) will improve. This is due to

the *multi-user diversity* (MUD) effect: when there are many users with independently fading channels, at any one time there is a high probability that one of the users will have a strong channel. By allowing only that user to transmit, the resources are used in the most efficient manner and the total system throughput is maximized. The larger the number of users, the stronger tends to be the strongest channel, and the more the multi-user diversity gain [Tse and Viswanath, 2005]. This MUD is inherent in the system and we do not need to create it in contrast to time, frequency or space diversity. The MUD gain arises due to improvement in the channel gain from $g_i(n)$ to $\max_i g_i(n)$ in time-slot n with N users in the system. However, always selecting the user with the best channel conditions can result in the starvation of other users, since channel conditions often change slowly. To avoid such starvation, it is important to consider different QoS demands of the users in addition to the channel conditions when designing scheduling algorithms.

Opportunistic scheduling requires the availability of the CSI at the base station. However, this information is already available from the link adaptation. If it is assumed that all the users' CSI is available for each time-slot, the base station can use this information to perform scheduling on a per time-slot basis.

Some of the well-known opportunistic scheduling algorithms are discussed below:

1.6.1 Maximum Carrier-to-Noise Ratio Scheduling

The *maximum carrier-to-noise ratio scheduling* (MCS) algorithm is the simplest opportunistic scheduling algorithm. In time-slot n, the MCS algorithm schedules the user with the highest CNR [Knopp and Humblet, 1995]:

$$i^*(n) = \underset{1 \le i \le N}{\operatorname{argmax}} \ \gamma_i(n), \tag{1.10}$$

where $\gamma_i(n)$ is the CNR of user i in time-slot n. The MCS algorithm maximally exploits the MUD in a time-slotted system where only one user is scheduled at a time. However, this algorithm is not fair and may result in the starvation of other users, as discussed above.

1.6.2 Normalized Carrier-to-Noise Ratio Scheduling

The *normalized carrier-to-noise ratio scheduling* (NCS) algorithm schedules the user that has the highest ratio of instantaneous to average CNR in the time-slot n [Yang and Alouini, 2006]:

$$i^*(n) = \underset{1 \le i \le N}{\operatorname{argmax}} \frac{\gamma_i(n)}{\bar{\gamma}_i}, \tag{1.11}$$

where $\bar{\gamma}_i$ is the average CNR of user i. The NCS algorithm increases the fairness of the system.

1.6.3 Proportional Fair Scheduling

The *proportional fair scheduling* (PFS) algorithm is described in [Chaponniere, Black, Holtzman, and Tse, 2002; Holtzman, 2001; Viswanath, Tse, and Laroia, 2002]. The base station keeps track of the average throughput $\mathcal{T}_i$ of each user in an exponentially weighted time-window of length t_c [seconds]. Before making a scheduling decision in time-slot n, the base station receives the requested rates $r_i(n)$s from all the users, and schedules the user $i^*(n)$ as follows:

$$i^*(n) = \underset{1 \le i \le N}{\operatorname{argmax}} \left(\frac{r_i(n)}{\mathcal{T}_i(n)} \right) \tag{1.12}$$

For each user i, the average throughput $\mathcal{T}_i(n)$ is updated with an exponentially weighted factor, and is given by

$$\mathcal{T}_i(n+1) = \begin{cases} \left(1 - \frac{1}{t_c}\right) \mathcal{T}_i(n) + \frac{1}{t_c} r_i(n), & i = i^*(n) \\ \left(1 - \frac{1}{t_c}\right) \mathcal{T}_i(n), & i \neq i^*(n) \end{cases} \tag{1.13}$$

When there are many users in a cell, this algorithm ensures both that the users are scheduled close to their own peak CNR and that they have the same probability of being scheduled in a randomly picked time-slot [Avidor, Mukherjee, Ling, and Papadias, 2004].

1.6.4 Opportunistic Round Robin Scheduling

In [Kulkarni and Rosenberg, 2003], the idea of *opportunistic round robin* (ORR) algorithm was first proposed. This algorithm is stated as an optimization problem where the throughput over a time-window of $K = N$

time-slots has to be maximized, subject to the constraint that all the N users should get exactly one time-slot each within the time-window.

A practical version of this algorithm is described in [Johansson, 2004], where the N users are scheduled in a round of $K = N$ time-slots as follows: In the first time-slot, the user with the highest CNR is selected. This user is then not considered in the remaining $N - 1$ time-slots. For the next time-slot the user with the highest CNR of the remaining users is scheduled. This procedure is repeated until the Nth time-slot, where the last remaining user is scheduled. The scheduling process then starts all over again with a new round of N time-slots.

1.6.5 Normalized Opportunistic Round Robin Scheduling

If the average CNRs of the users are spread far apart, the ORR algorithm will become non-opportunistic, i.e., the user with the highest average CNR will always be scheduled in the first time-slot of a round, the user with the second-highest average CNR will be scheduled in the second time-slot, and so on. In such a scenario, it is better to combine the ORR algorithm with the NCS algorithm to yield a higher system spectral efficiency. This algorithm is called the *normalized opportunistic round robin* (NORR) scheduling algorithm [Ji, Yang, Zhou, M.Takai, and Bagrodia, 2004]. In each time-slot n, the user with the highest ratio $\gamma_i(n)/\bar{\gamma}_i$ is scheduled. The rest of the procedure of NORR scheduling algorithm is same as that of ORR algorithm.

1.6.6 Modified Largest Weighted Delay First Scheduling

In [Andrews, Kumaran, Ramanan, Stolyar, Vijayakumar, and Whiting, 2000; Andrews, Kumaran, Ramanan, Stolyar, Whiting, and Vijayakumar, 2001], a modified version of the LWDF algorithm is proposed which takes both the user's channel conditions and the delay of packets into account. This is called *modified largest weighted delay first* (MLWDF) scheduling algorithm. The MLWDF algorithm schedules a user which maximizes

$$i^*(n) = \underset{1 \le i \le N}{\operatorname{argmax}} \left(\phi_i W_i(n) \frac{r_i(n)}{\bar{R}_i} \right), \tag{1.14}$$

where $W_i(n)$ and ϕ_i are defined above, and $\bar{R}_i$ [bits/second] is the average rate for user i.

1.7 Scheduling Multiple Users per Time-Slot

In the previous sections, scheduling algorithms for scheduling single user in each time-slot were discussed. However, in modern wireless networks it is often possible to schedule more than one user in each time-slot. In this section, we discuss some techniques that enable the selection of multiple users in a single time-slot.

1.7.1 MIMO Systems

Multi-user multiple-input multiple-output (MIMO) systems can be divided into two categories: MIMO broadcast channels (MIMO BC) and MIMO multiple access channels (MIMO MAC) for downlink and uplink transmissions, respectively.

In multi-user MIMO systems, the achievable capacity gains are limited by the number of receive antennas at each user. In the case of a single receive antenna, the base station should send data to multiple users simultaneously in order to achieve the capacity gains [Jagannathan, Borst, Whiting, and Modiano, 2006]. It is shown in [Jindal and Goldsmith, 2005] that the capacity gain over a MIMO system where only the user with the best channel is scheduled in each time-slot, is approximately $\min(M, N)$, where M is the number of transmit antennas at the base station, and N is the number of mobile users. The Shannon capacity of such systems can be obtained by using *dirty paper coding* (DPC) [Caire and Shamai, 2003; Viswanath and Tse, 2003]. However, DPC is difficult to implement in practical systems. It has been shown that systems using *beamforming*, where a group of users with the best semi-orthogonal channels are scheduled in each time-slot, can come close to the performance of DPC [Sharif and Hassibi, 2005].

1.7.2 OFDM/OFDMA Systems

With *orthogonal frequency division multiplexing* (OFDM), it is also possible for more users to transmit or receive within a time-slot. In OFDM, the channel bandwidth is divided into multiple sub-carriers, and each sub-carrier in each time-slot can be allocated to different users [Lawrey, 1999; Wong, Cheng, Letaief, and Murch, 1999]. Scheduling sub-carriers in this way means that the MUD can be further exploited and that it is easier to fulfill QoS guarantees within a time-window. Various scheduling algo-

rithms for multi-user OFDM (also known as OFDMA) have been proposed in literature [Conte, Tomasin, and Benvenuto, 2008; Girici, Zhu, Agre, and Ephremides, 2010; Nguyen and Han, 2006; Zhang, He, and Chong, 2005].

Allocating different sub-carriers to different users, however, requires that the signals for/from all the users are time and frequency synchronized to each other. Analysis of the impact of the time and frequency synchronization errors on the system performance is carried out in [Mostofi and Cox, 2006; Stemick and Rohling, 2007].

1.7.3 CDMA Systems

CDMA systems also allow multiple users to transmit or receive data in a single time-slot. Each user is assigned a code that is different from the codes of other users that share the same radio channel. However, every user will receive interference from every other user due to non-zero cross-correlations between different users' codes. Moreover in the uplink, signals originating from different users will arrive at the base station with unequal power levels because of different locations within the cell. A distant user whose received signal at the base station is low will suffer due to the interference from the nearby user whose received signal level is high. This is known as the *near-far* problem. Power control is therefore important to reduce the interference in a CDMA-based network [Maruddani and Kurniawan, 2010; Sung and Wong, 2001]. Moreover, the throughput can be increased further by using superposition coding with successive interference cancellation (SCSIC) [Madkour, Gupta, and Wang, 2002].

1.8 Performance Criteria for Scheduling Algorithms

In the following, we discuss some possible performance criteria for the scheduling algorithms.

1.8.1 Throughput Guarantees

With real-time traffic transmitted over wireless networks, the need for more exact QoS measures is in the interests of both network operators and customers. The customers want to know what they have bought and the operators would rather not give away more network capacity to the customers

than they have paid for. A measure that is well suited to quantify QoS guarantees exactly is a *throughput guarantee*, which is defined as the number of bits B, a user is guaranteed to transmit or receive within a time-window T_W [seconds]. Throughput guarantees can in principle be either *hard* (*deterministic*) or *soft* (*statistical*) [Ferrari, 1990]:

- Hard throughput guarantees promise with unit probability that a guarantee will be fulfilled over a time-window T_W, i.e.,

$$b \geq B, \tag{1.15}$$

  where b is the actual number of bits provided by the system over the time-window T_W.

- The soft throughput guarantees promise with a lower than unity – but preferably high – probability that the specified throughput guarantee will be fulfilled, i.e.,

$$\Pr[b \geq B] \geq \zeta_B, \tag{1.16}$$

  where ζ_B is a lower bound on the probability that the system will provide a throughput greater than the one promised to the user. The parameter ζ_B should be as close to unity as possible.

For telecommunication networks in general, and for wireless networks in particular, soft throughput guarantees are more suitable for specifying QoS than hard throughput guarantees. This is due to the varying number of users, varying loads from the applications of these users, and the varying quality of the radio channel.

Definition 1.1 (Throughput Guarantee Violation Probability (TGVP)) *The probability of not fulfilling a throughput guarantee B [bits] within a specified time-window T_W [seconds], averaged over all N users in the system [Hassel, Øien, and Gesbert, 2007].*

For a specific user i, the TGVP_i is the probability of the number of bits b_i transmitted to or from it within a time-window T_W being below B_i, and is denoted as:

$$\text{TGVP}_i = \Pr[b_i < B_i],\ i = 1, 2, \cdots, N. \tag{1.17}$$

The TGVP for the overall system is then given as

$$\text{TGVP} = \frac{1}{N} \sum_{i=1}^{N} \text{TGVP}_i. \tag{1.18}$$

Throughput guarantees in most cases cannot be given with absolute certainty (i.e., we focus on *soft* throughput guarantees). It is however important that the guaranteed number of bits B_i within the time-window T_W should be promised with high probability. This means that when assessing the relative behavior of different scheduling algorithms, the TGVP performance of the algorithms close to TGVP $\cong 0$ is the most interesting [Hassel, de la Kethulle de Ryhove, and Øien, 2007; Rasool, Hassel, de la Kethulle de Ryhove, and Øien, 2011].

1.8.2 Average System Spectral Efficiency

The *Average System Spectral Efficiency* (ASSE) is an important performance measure when designing the scheduling algorithms. ASSE is defined as the theoretically attainable throughput per bandwidth [bits/s/Hz], averaged over all the N users in the system. The expression for the ASSE for constant-power, optimal rate adaptation, assuming AWGN noise, is defined as [Goldsmith and Varaiya, 1997]:

$$\text{ASSE} = \sum_{i=1}^{N} p(i) \int_0^{\infty} \log_2(1+\gamma) f_{\gamma^*}(\gamma|i)\, d\gamma, \tag{1.19}$$

where $p(i)$ is the access probability of user i, and $f_{\gamma^*}(\gamma|i)$ is the PDF of the CNR of user i when this user is selected. This definition of ASSE assumes that only one user is scheduled at every time-slot.

1.8.3 Fairness

Fairness is a measure of how equally the radio resources are allocated among the mobile users. A commonly used fairness measure is *Jain's Fairness Index* (JFI) [Hassel, Hanssen, and Øien, 2006; Jain, Chiu, and Hawe, 1984]:

$$\text{JFI} = \frac{(\mathbb{E}_K[X])^2}{\mathbb{E}_K[X^2]}, \tag{1.20}$$

where X is a random variable describing the amount of resource allocated to a user, and $\mathbb{E}_K[\cdot]$ is the expectation calculated over the distribution of the resource allocation over K time-slots.

The scheduling algorithms that achieve high fairness over a relatively short time-window are termed as *short-term fair*, while the algorithms that obtain high fairness over an infinitely long time-window are called *asymptotically fair* [Hassel et al., 2006].

1.8.4 Delay

Delay can be defined as the time from the service provider sending the data to the user receiving it. The delay requirements can also be *hard* or *soft* [Ferrari, 1990]:

- Hard Delay Guarantee:

$$d \leq D, \tag{1.21}$$

  where d is the delay with which the data sent by the service provider is received by the user, and D is the maximum delay that is promised.

- Soft Delay Guarantee:

$$\Pr[d \leq D] \geq \zeta_D, \tag{1.22}$$

  where ζ_D is a lower bound on the probability of timely data delivery.

Real-time applications such as VoIP and video conferencing have strict delay requirements. Outdated packets cannot be used and are discarded by the system. Higher priorities can be given to the applications that have stringent delay requirements.

1.8.5 Power/Energy Consumption

The performance criterion for some networks is low power/energy consumption [Miao, Himayat, Li, and Swami, 2009]. Wireless sensor networks is an example of such networks. Since sensors typically run on small batteries, management of the available energy directly impacts the sensor network's operation lifetime and performance. Therefore, the schemes that optimize the sensor's energy consumption have great importance [Cardei, Thai, Li, and Wu, 2005].

1.9 Literature Survey

In this section, we review some articles that focus on providing throughput guarantees to the users in the network.

As mentioned earlier, opportunistic multi-user scheduling gives higher throughput in a wireless cell than non-opportunistic algorithms like Round Robin algorithm. However, always selecting the users with the best channel quality may lead to starvation of other users. A better approach is to

have a fair resource allocation among the users, e.g., by using PFS algorithm.

A general framework for opportunistic scheduling is presented in [Liu, Chong, and Shroff, 2003], along with three general categories of scheduling problems under this framework. The third category, i.e. *minimum performance requirement* discusses the scenario that is similar to the scenarios considered in this dissertation. A stochastic-approximation-based algorithm is also provided to estimate the key parameters of the scheduling scheme online. However, the merit and novelty of our work is that our scheduling algorithms are significantly simpler and thus more applicable than the one proposed in [Liu et al., 2003].

In [Andrews, Qian, and Stolyar, 2005], Andrews et al. propose scheduling algorithms that aim at fulfilling throughput guarantees by giving different priorities to the users depending on how far they are from their maximum and minimum throughput guarantees. One of the problems with this algorithm is that it takes action only when a throughput guarantee has been violated. The authors have therefore shown in [Andrews et al., 2005], how time parameters of their algorithm can be set shorter than the actual time-window of interest to alleviate this issue. The scheduling algorithms proposed in this thesis try to fulfill the throughput guarantees before they are violated.

A utility-based predictive scheduler is proposed in [Redana, Frediani, and Capone, 2008] that focuses on fulfilling the throughput guarantees by predicting the future channel conditions and adopting the rates accordingly. At the current time-slot, it schedules the user whose future channel conditions would make it more difficult to provide the throughput guarantees.

Borst and Whiting have elegantly proved that a certain scheduling policy provides the highest throughput guarantee for wireless networks [Borst and Whiting, 2003]. However, they briefly argue that the rate distributions of the users are unknown and they have therefore not shown how this optimal scheduling policy can be found for users with differently distributed CNRs.

While much research has been done on providing *long-term* throughput guarantees, little work has addressed how to guarantee the *short-term* throughput to the users. The research in short-term performance has mainly focused on fairness issues. Our proposed adaptive scheduling algorithms

are thus significant. We now discuss some works that focus on short-term throughput guarantees specifically.

In [S. Lu and Srikant, 1999], the authors extend wireline scheduling policies to wireless networks and present wireless fair scheduling policies which give short-term and long-term throughput guarantee bounds.

The authors in [Kastrinogiannis and Papavassiliou, 2007], have analyzed and evaluated the problem of real-time users' short-term QoS probabilistic properties, in terms of maximum delay and minimum received throughput guarantees, under basic opportunistic scheduling policies (MCS and PFS). In [Kastrinogiannis and Papavassiliou, 2008], the authors argued that the probabilistic delay constraints are insufficient indicators of real-time QoS requirements, and probabilistic short-term throughput guarantees are more appropriate criteria. Based on this argument, they developed and evaluated a utility based opportunistic resource allocation algorithm which aims at the minimization of real-time users' short-term TGVPs.

An algorithm aimed specifically at providing short-term throughput guarantees has been proposed in [Chen and Jordan, 2009]. However, this algorithm achieves a significantly lower average long-term throughput as compared to MCS or PFS algorithm.

In [Bang, Ekman, and Gesbert, 2008], the authors proposed a predictive proportional fair algorithm and showed that its short-term throughput performance is better than the PFS algorithm.

1.10 Outline of the Thesis

The rest of the dissertation is organized as follows:

Chapter 2: In Chapter 2, we develop an expression for the approximate TGVP for users in time-slotted networks with the given cumulants of the distribution of bit-rate in a time-slot, and a given distribution for the number of time-slots allocated within a time-window. Through simulations, it is shown that this TGVP approximation is tight for a realistic wireless network with moving users and correlated channels.

This chapter is based on [Rasool et al., 2011; Rasool and Øien, 2011a].

Chapter 3: We formulate an optimization problem which aims at maximizing the throughput that can be guaranteed to the mobile users in Chapter 3. By building on results obtained by Borst and Whiting

and by assuming that the distributions of the users' CNRs are known, we find the solution to this problem for users with different channel quality distributions, for both the scenario where all the users have the same throughput guarantees, and the scenario where all the users have different throughput guarantees. Based on these solutions, we also propose two simple and low complexity adaptive scheduling algorithms that perform significantly better than other well-known scheduling algorithms. We also analyze the ASSE and fairness of the proposed optimal scheduling algorithm in this chapter.

This chapter is based on [Hassel et al., 2007; Rasool et al., 2011; Rasool and Øien, 2010a,b].

Chapter 4: In Chapter 4, we assume that a maximum a posteriori (MAP) predictor is employed for the CNR prediction, so that the system takes the feedback delay and the channel noise into account. We then investigate the effect of imperfect channel prediction and delay on the throughput guarantees promised to all the users in the wireless network. A procedure to reduce the probability of outage in case of imperfect channel prediction is also proposed.

This chapter is based on [Rasool and Øien, 2012b].

Chapter 5: We formulate an optimization problem that aims at maximizing the throughput guarantees offered in a MIMO broadcast channel in Chapter 5. We also propose two scheduling algorithms that make use of orthonormal random beamforming, and try to fulfill the throughput guarantees promised to all the mobile users. The scheduling algorithms are designed for two different beamforming scenarios. In the first scenario, random beams are used for all the selected users, whereas transmit beamforming is used for the first selected user in the second scenario.

This chapter is based on [Rasool and Øien, 2011b,c].

Chapter 6: In Chapter 6, we consider the issue of discrete power and bandwidth allocation for discrete-rate multi-user link adaptation with imperfect channel state information. To be more specific, we discuss how the ACM scheme can be designed in such a scenario for i) sum rate maximization and ii) average power minimization in a multi-user setting.

This chapter is based on [Rasool and Øien, 2012a].

Chapter 7: We conclude with the main contributions of this research work in Chapter 7. We also discuss some of the open problems and their impact on the improvement of proposed scheduling schemes.

Appendix A: In Appendix A, we show how to obtain the set T_i^N by expanding the product in (2.16).

Appendix B: In this appendix, we derive the expression for $\Psi(x)$ that is used in (2.21).

Appendix C: In Appendix C, we analyze the correctness of the parameters of the TGVP expression for the MTGS algorithm, derived in Section 3.6.

Appendix D: The expressions for the ASSE of the RR, MCS and NCS algorithms is provided in Appendix D.

Appendix E: In Appendix E, the expressions for the time-slot and throughput fairness for the RR, MCS and NCS algorithms is provided.

Appendix F: We derive (4.14) in Appendix F.

1.11 Papers Not Included in the Thesis

In addition to the publications that this thesis is based on, the author has also contributed to the following papers, as mentioned briefly in the following:

- J. Rasool, G. E. Øien, J. E. Håkegård and T. A. Myrvoll "On multiuser MIMO capacity benefits in air-to-ground communication for airport traffic management," in *Proc. International Symposium on Wireless Communication Systems (ISWCS' 09)*, Siena, Italy, pp. 458 - 462, Sept 2009. [Rasool, Øien, Håkegård, and Myrvoll, 2009]

- J. Rasool and G. E. Øien, "Multiuser MIMO systems for spectrally efficient communications in future aeronautical services," in *VERDIKT Conference*, Oslo, Norway, Nov 2009. [Rasool and Øien, 2009]

  There is an urgent need to increase the capacity of air-to-ground communication systems for future aeronautical communication services.

In these papers, we have investigated the utilization of multi-user MIMO communication strategies in aeronautical communication, since it can offer high capacity gains. We model airplane-Air Traffic Control (ATC) communication architecture on a MIMO MAC system. Our focus is primarily on LoS channels since it is the most appropriate channel model for most air-to-ground communications except for communications with airplanes located at the gate. We model the antennas at the ATC as a uniform linear array (ULA) and quantify the potential capacity performance of MIMO MAC systems when employed in future air-to-ground communication scenarios. As a by-product we also determine the appropriate spacing between the antennas to be deployed at the ATC tower, so that the maximal achievable capacity gains are realized. This spacing is an important parameter when deciding whether or not such systems are implementable in practice. Since we operate at high frequencies (1 GHz corresponding to L-band), it turns out that the required antenna separation is of the order of a few centimeters. This will allow us to use ULAs with many antennas at the ATC tower, thus realizing large potential capacity gains. Indeed, if we want to employ a very large number of antennas, we can use the whole terminal building for that purpose. This would allow us to utilize MIMO MAC communication to efficiently improve the system capacity, which is sorely needed for future aeronautical services.

Chapter 2

Quantifying the Offered Throughput Guarantees

As mentioned before, the search for more exact QoS measures is in the interests of both the network operators and the customers. This is more true for real-time traffic transmitted over wireless networks. The customers want to know what they have bought, and the operators would rather not give away more network capacity to the customers than they have paid for. A measure that is well suited to quantify QoS guarantees exactly is a *throughput guarantee*, as defined in Section 1.8.1. In real-world wireless networks, it is valuable for the network providers if they are somehow able to quantify the (soft) throughput guarantees for a certain scheduling algorithm without conducting experimental investigations. The advantage of this quantification is two-fold; first, it will make it easier for them to offer a service that is tailor-made to the user applications that are to be transmitted. Secondly, they do not have to over-dimension their wireless networks to fulfill the QoS demands of the users.

In this chapter, we develop an expression for the TGVP (see Section 1.8.1) for users in time-slotted networks, for any scheduling algorithm, under the assumption of given cumulants (at least the first two cumulants corresponding to the mean and variance) of the distribution of the bit-rate in a time-slot, and a given distribution for the number of time-slots allocated within a time-window. We shall also show that our TGVP approximation is accurate for a wireless network, with fast moving users having correlated channels.

In [Hassel et al., 2007], the authors have also derived an approximate expression for TGVP by using the *central limit theorem*. Although their expression in practice provides a very good TGVP approximation, we argue that since the users are generally offered (soft) throughput guarantees with close to unit probability, the probability of violating a throughput guarantee is very small, i.e. close to zero. Therefore, violation of the throughput guarantee should be treated as a *rare* event. *Large deviation theory* (LDT) is a branch of probability theory that deals with rare events, and provides asymptotic estimates for their probabilities. We shall use Cramér's theorem from LDT to derive a novel approximate TGVP expression.

The rest of the chapter is organized as follows. In Section 2.1 we present the system model, and in Section 2.2, we derive the expression for the approximate TGVP. In Section 2.3, we derive the equations for the necessary parameters of the TGVP expression for MCS and NCS algorithms. We provide the system/simulation parameters in Section 2.4, and present our numerical results in Section 2.5. A brief summary of the chapter is given in Section 2.6.

This chapter is based on [Rasool et al., 2011; Rasool and Øien, 2011a].

2.1 System Model

We consider a single base station that serves N backlogged users using time-division multiplexing (TDM). The analysis conducted in this chapter is valid both for the uplink and the downlink; in either case we assume that the total available bandwidth for the users is W [Hz] and that the users have constant transmit power. Each user estimates his CNR perfectly, and it is assumed that the base station receives these measurements from all the users before downlink scheduling is performed. Uplink scheduling for each time slot is also performed by the base station based on perfect channel knowledge, and the scheduling decision is distributed to the selected user before uplink transmission starts.

The communication channel between the base station and all the users is modeled by a flat, Rayleigh block-fading channel subject to AWGN noise; and it is assumed that the communication channels corresponding to the different users fade independently. The duration of a block is taken as one time-slot, and is denoted as T_{TS} [seconds]. To obtain our analytical results we assume that the CNR values from time-slot to time-slot are uncorre-

lated. However, for simulations, we assume that the CNR values corresponding to different time-slots are correlated. The correlation model used in our simulations is described in Section 1.3.2.

The average CNR of user i, denoted by $\bar{\gamma}_i$, is assumed to be constant in the time-window over which the throughput guarantees are calculated. This can be realistic for a real-life wireless network because the average CNR of the users' CNR distributions normally changes on a time-scale of several seconds, while the throughput guarantees are often calculated over time-windows of less than one hundred milliseconds [Hassel et al., 2007].

The probability distributions of the CNRs of all the users are assumed to be known perfectly at the base station. In modern cellular standards like HSDPA, LTE and Mobile WiMAX [WiMAX Forum, 2006], much of the information needed for obtaining precise probability distribution estimates is already available. For link adaptation, modern cellular networks have precise, real-time CNR estimates of the users. These channel quality estimates can also be used to obtain estimates of the probability distributions of the CNRs of each one of the users.

It is also assumed that the population of backlogged users is constant and equal to N. This is a realistic assumption since the separation of time-scales makes the population of backlogged users nearly static; i.e., the population of backlogged users changes much slower than the time-window over which the throughput guarantees are calculated [Borst and Whiting, 2003].

We further assume that the users always have data to send. This is a realistic assumption for real-time applications because the packet flow from such applications is relatively constant [Hassel et al., 2007].

Finally, it is also assumed that only one user can be scheduled in a time-slot.

2.2 Derivation of (Approximate) TGVP Expression

In this section, we derive an expression for the approximate TGVP, that can be used to specify achievable soft throughput guarantees in the network. It is assumed that the data is being transmitted over a time-slotted block fading channel, and user i is promised a throughput guarantee of B_i bits over a time-window T_W constituting K time-slots.

Within a time-window consisting of K time-slots, a user will get either a total of $0, 1, 2, \cdots, K-1$ or K time-slots, but no two of these concurrently. Therefore the allocation of different number of time-slots to a user i constitutes mutually exclusive events. Using the law of total probability, the TGVP for user i, given in (1.17), can thus be expressed as follows:

$$\begin{aligned} \text{TGVP}_i &= \Pr[b_i < B_i] \\ &= \Pr[b_i < B_i|0] \cdot p_K(0|i) \\ &+ \Pr[b_i < B_i|1] \cdot p_K(1|i) \\ &... \\ &+ \Pr[b_i < B_i|K] \cdot p_K(K|i), \end{aligned} \tag{2.1}$$

where $\Pr[b_i < B_i|k]$ denotes the TGVP when user i is allocated k time-slots and $p_K(k|i)$ is the probability that user i gets k time-slots out of K time-slots.

We denote the number of bits transmitted to or from user i within the jth time-slot he is scheduled by $b_{i,j}$. We consider a system having constant transmit power and capacity-achieving codes, which operate at the Shannon capacity limit. $b_{i,j}$ is then given as

$$b_{i,j} = T_{\text{TS}} W \log_2(1 + \gamma_{i,j}),$$

where $\gamma_{i,j}$ is the CNR in the jth time-slot user i is scheduled, and T_{TS} is the length of the time-slot in seconds. The mean and variance of $b_{i,j}$ are denoted by $\mu_{b_{i,j}}$ and $\sigma^2_{b_{i,j}}$ respectively, and are given as follows:

$$\mu_{b_{i,j}} = m_1 = \mathbb{E}[b_{i,j}], \tag{2.2}$$

and

$$\sigma^2_{b_{i,j}} = m_2 - m_1^2 = \mathbb{E}[b^2_{i,j}] - (\mathbb{E}[b_{i,j}])^2, \tag{2.3}$$

where m_l is the lth order moment of the distribution of $b_{i,j}$.

The probability for violating the throughput guarantee B_i when k out of K time-slots are allocated to the user i can be given as:

$$\begin{aligned} \Pr[b_i < B_i|k] &= \Pr\left[\sum_{j=1}^{k} b_{i,j} < B_i\right] \\ &= \Pr\left[\bar{b}_{i,k} < \frac{B_i}{k}\right], \end{aligned} \tag{2.4}$$

where

$$\bar{b}_{i,k} = \frac{1}{k}\sum_{j=1}^{k} b_{i,j}$$

is the average number of bits being transmitted to or from user i when he is allocated k time-slots, and we assume that $\mu_{\bar{b}_{i,k}}$ and $\sigma^2_{\bar{b}_{i,k}}$ are the mean and variance of $\bar{b}_{i,k}$, respectively.

Since the users are generally offered (soft) throughput guarantees with close to unit probability, the probability of violating a throughput guarantee should be very small, i.e. close to zero. In this derivation, we shall therefore treat the violation of the throughput guarantee as a *rare* event. As stated before, LDT is a branch of probability theory that deals with rare events and provides asymptotic estimates for their probabilities. We shall use Cramér's theorem from LDT [Bucklew, 2004, p. 27] to derive the approximate TGVP expression in what follows.

We need to consider the following two cases when applying Cramér's theorem: For $\frac{B_i}{k} < \mu_{b_{i,j}}$, we have

$$\lim_{k\to\infty} \frac{1}{k} \ln \Pr\left[\bar{b}_{i,k} \leq \frac{B_i}{k}\right] = -I(\frac{B_i}{k}),$$

where $I(\cdot)$ is known as the *Cramér function* or *large deviation rate-function*[1] [Bucklew, 2004, p. 28]. The function $I(\cdot)$ is defined as the Legendre-Fenchel transform [Hugo Touchette, 2007] of the cumulant generating function $\lambda(\theta)$.

$$I\left(\frac{B_i}{k}\right) = \sup_{\theta}\left(\theta\frac{B_i}{k} - \lambda(\theta)\right) \tag{2.5}$$

The cumulant generating function $\lambda(\theta)$ is the natural logarithm of the moment generating function $M(\theta)$ and its Taylor expansion is given as follows [Eric W. Weisstein, b]:

$$\begin{aligned}\lambda(\theta) &= \ln M(\theta) \\ &= \kappa_1\theta + \kappa_2\frac{\theta^2}{2!} + \kappa_3\frac{\theta^3}{3!} + ...,\end{aligned}$$

[1] It is called rate-function because it gives the rate at which the tail of the probability distribution decays exponentially with increasing k.

where $\kappa_1, \kappa_2, \kappa_3, ...$ are the cumulants, which can be calculated from the moments of the distribution of $b_{i,j}$ as follows [Eric W. Weisstein, a]:

$$\begin{aligned} \kappa_1 &= m_1 = \mu_{b_{i,j}} \\ \kappa_2 &= m_2 - m_1^2 = \sigma^2_{b_{i,j}} \\ \kappa_3 &= m_3 - 3m_2m_1 + 2m_1^3 \\ &\vdots \end{aligned}$$

From this, it follows that

$$\Pr\left[\bar{b}_{i,k} < \frac{B_i}{k}\right] \approx e^{-kI(B_i/k)}. \tag{2.6}$$

Furthermore, for $\frac{B_i}{k} > \mu_{b_{i,j}}$, we get

$$\begin{aligned} \lim_{k\to\infty} \frac{1}{k} \ln \Pr\left[\bar{b}_{i,k} \geq \frac{B_i}{k}\right] &= -I(\frac{B_i}{k}) \\ \Rightarrow \Pr\left[\bar{b}_{i,k} \geq \frac{B_i}{k}\right] &\approx e^{-kI(B_i/k)}, \end{aligned}$$

$$\Rightarrow \Pr\left[\bar{b}_{i,k} < \frac{B_i}{k}\right] \approx 1 - e^{-kI(B_i/k)}. \tag{2.7}$$

In this work, we only consider the first two cumulants for simplification. However, we must emphasize that higher order cumulants should also be included for more accurate results. The cumulant generating function when considering the first two cumulants is then given as

$$\lambda(\theta) = \theta\mu_{b_{i,j}} + \frac{\sigma^2_{b_{i,j}}}{2}\theta^2. \tag{2.8}$$

Substituting (2.8) in (2.5),

$$I\left(\frac{B_i}{k}\right) = \sup_{\theta}\left(\theta\frac{B_i}{k} - \theta\mu_{b_{i,j}} - \frac{\sigma^2_{b_{i,j}}}{2}\theta^2\right). \tag{2.9}$$

The value of θ^* that maximizes (2.9) is found to be

$$\theta^* = \frac{\frac{B_i}{k} - \mu_{b_{i,j}}}{\sigma^2_{b_{i,j}}}. \tag{2.10}$$

Thus the large deviation rate-function in this case is given as

$$I\left(\frac{B_i}{k}\right) = \frac{\left(\frac{B_i}{k} - \mu_{b_{i,j}}\right)^2}{2\sigma^2_{b_{i,j}}}. \tag{2.11}$$

Finally, the probability that the throughput constraint B_i is violated over K time-slots for user i can be approximated as:

$$\Pr[b_i < B_i] \approx p_K(0|i) + \sum_{k=1}^{K} p_K(k|i)\Pr[b_i < B_i|k], \tag{2.12}$$

where $\Pr[b_i < B_i|k]$ is given in (2.6) and (2.7) for the two cases discussed.

The TGVP for the overall system is then given as

$$\text{TGVP} = \frac{1}{N}\sum_{i=1}^{N}\Pr[b_i < B_i]. \tag{2.13}$$

Through simulations (Section 2.5), we will show that our TGVP approximation is tight for a realistic network with moving users and correlated channels.

2.3 Scheduling Schemes for TGVP Verification

We use the MCS and NCS algorithms to verify the proposed TGVP expression. In this section, we derive the equations for the necessary parameters of the TGVP expression for these two schemes.

2.3.1 MCS Algorithm

According to [Hassel et al., 2007], for the MCS scheme, the number of time-slots allocated to a user i within a time-window of K time-slots is distributed according to the *binomial distribution* [Kreyszig, 1993, p. 1179]:

$$p_K(k|i) = \binom{K}{k} p(i)^k (1 - p(i))^{K-k}, \tag{2.14}$$

where $p(i)$ is the probability of selecting user i in a time-slot. It can be expressed as [Yang and Alouini, 2006, Eq. (13)]:

$$\begin{aligned} p(i) &= \int_0^\infty f_{\gamma_i}(\gamma) \prod_{\substack{j=1 \\ j\neq i}}^{N} F_{\gamma_j}(\gamma) d\gamma \\ &= \frac{1}{\bar{\gamma}_i} \sum_{\tau \in T_i^N} \text{sign}(\tau) \frac{1}{\frac{1}{\bar{\gamma}_i} + \tau}, \end{aligned} \tag{2.15}$$

where $f_{\gamma_i}(\gamma)$ and $F_{\gamma_j}(\gamma)$ are the probability density function (PDF) and the cumulative distribution function (CDF) of the CNR of a single user with average CNR $\bar{\gamma}_i$ and $\bar{\gamma}_j$ respectively. T_i^N denotes the set containing the terms obtained by expanding the product

$$\prod_{\substack{j=1 \\ j\neq i}}^{N} F_{\gamma_j}(\gamma) = \prod_{\substack{j=1 \\ j\neq i}}^{N} (1 - e^{-\gamma/\bar{\gamma}_j}) \tag{2.16}$$

and then taking $(-1/\gamma)$ times the natural logarithm of the absolute value of each term independently, and $\text{sign}(\tau)$ is the corresponding sign of each term in the expansion [Eng, Kong, and Milstein, 1996, Sec. III-D2],[Alouini and Simon, 2000, Sec. VI-A2]. An example to obtain the set T_i^N is given in Appendix A. It is also assumed that the users have Rayleigh distributed channels.

For the MCS scheme, the PDF of the CNR of user i when this user is scheduled for transmission/reception, denoted by $f_{\gamma*}(\gamma|i)$, is given as [Hassel, 2007]

$$f_{\gamma*}(\gamma|i) = \frac{f_{\gamma_i}(\gamma)}{p(i)} \prod_{\substack{j=1 \\ j\neq i}}^{N} F_{\gamma_j}(\gamma). \tag{2.17}$$

The first moment m_1 (the mean value for $b_{i,j}$) for the MCS scheme can be derived as follows [Yang and Alouini, 2006]:

$$\begin{aligned} m_1 &= \mathbb{E}[b_{i,j}] \\ &= WT_{\text{TS}} \int_0^\infty \log_2(1+\gamma) f_{\gamma*}(\gamma|i) d\gamma \\ &= \frac{WT_{\text{TS}}}{p(i)\bar{\gamma}_i \ln 2} \sum_{\tau \in T_i^N} \text{sign}(\tau) \frac{e^{\left(\frac{1}{\bar{\gamma}_i} + \tau\right)}}{\frac{1}{\bar{\gamma}_i} + \tau} E_1\left(\frac{1}{\bar{\gamma}_i} + \tau\right), \end{aligned} \tag{2.18}$$

where

$$E_1(x) = \int_1^\infty \frac{e^{-xt}}{t} dt \tag{2.19}$$

is the exponential integral function of first order. Similarly, the second moment m_2 of the number of bits $b_{i,j}$ transmitted to or from user i, in the jth time-slot he is scheduled, can be obtained as follows [Hassel et al., 2007]:

$$\begin{aligned} m_2 &= \mathbb{E}[b_{i,j}^2] \\ &= (WT_{\mathrm{TS}})^2 \int_0^\infty (\log_2(1+\gamma))^2 f_{\gamma*}(\gamma|i) d\gamma \\ &= \frac{(WT_{\mathrm{TS}})^2}{\bar{\gamma}_i p(i) (\ln 2)^2} \sum_{\tau \in T_i^N} \mathrm{sign}(\tau) \Psi\left(\frac{1}{\bar{\gamma}_i} + \tau\right), \end{aligned} \tag{2.20}$$

where $\Psi(x)$ is given by [Hassel et al., 2007]

$$\begin{aligned} \Psi(x) &= \int_0^\infty (\ln(1+\gamma))^2 e^{-x\gamma} d\gamma \\ &= e^x \left\{ \frac{1}{x} \left[\frac{\pi^2}{6} + (C_e + \ln(x))^2 \right] - 2_3F_3(1,1,1;2,2,2;-x) \right\}, \end{aligned} \tag{2.21}$$

where $C_e = 0.57721566490$ is Euler's constant, and ${}_pF_q(a_1,...,a_p;b_1,...,b_q;\cdot)$ is the generalized hypergeometric function [Eric W. Weisstein, c]. The derivation of $\Psi(x)$ is given in Appendix B.

2.3.2 NCS Algorithm

Since all the users in our system model have the same distribution for their relative CNRs [Yang and Alouini, 2006], and the *relatively* best user is scheduled in each time slot, the probability of scheduling a user in a time-slot is the same for all the users. Thus, for the NCS scheme, $p(i) = 1/N$ [Hassel et al., 2007]. The number of time-slots allocated to a user i within K time-slots is also distributed according to the binomial distribution expressed in (2.14). Therefore, we have

$$p_K(k|i) = \binom{K}{k} \frac{1}{N}^k (1 - \frac{1}{N})^{K-k}. \tag{2.22}$$

For the NCS scheme, the PDF of the CNR of user i when he is scheduled is given as [Hassel, 2007]

$$f_{\gamma*}(\gamma|i) = N F_{\gamma_i}^{N-1}(\gamma) f_{\gamma_i}(\gamma). \tag{2.23}$$

The first moment m_1 for the NCS scheme is given as follows [Alouini and Goldsmith, 1999, Eq. (44)]:

$$\begin{aligned} m_1 &= \mathbb{E}[b_{i,j}] \\ &= WT_{\mathrm{TS}} \int_0^\infty \log_2(1+\gamma) f_{\gamma*}(\gamma|i) d\gamma \\ &= \frac{NWT_{\mathrm{TS}}}{\ln 2} \sum_{j=0}^{N-1} \binom{N-1}{j} \frac{(-1)^j}{1+j} e^{\frac{1+j}{\bar{\gamma}_i}} E_1\left(\frac{1+j}{\bar{\gamma}_i}\right). \end{aligned} \tag{2.24}$$

The second moment m_2 of the number of bits $b_{i,j}$ transmitted to or from user i, in the jth time-slot he is scheduled, is given as follows [Hassel et al., 2007]:

$$\begin{aligned} m_2 &= \mathbb{E}[b_{i,j}^2] \\ &= (WT_{\mathrm{TS}})^2 \int_0^\infty (\log_2(1+\gamma))^2 f_{\gamma*}(\gamma|i) d\gamma \\ &= \frac{N(WT_{\mathrm{TS}})^2}{\bar{\gamma}_i (\ln 2)^2} \sum_{j=0}^{N-1} \binom{N-1}{j} (-1)^j \Psi\left(\frac{1+j}{\bar{\gamma}_i}\right). \end{aligned} \tag{2.25}$$

2.4 System/Simulation Parameters

For the Mobile WiMAX network, the duration of a time-slot for the downlink is 5 ms [WiMAX Forum, 2006]. The corresponding time-slot length for the 3GPP LTE network is 1 ms [Freescale Semiconductor, 2007]. The maximum acceptable delay of one way VoIP transmission must be under 80 ms and 50 ms for HSDPA and LTE systems respectively [K. Aho, Repo, Puttonen, Henttonen, Moisio, Kurjenniemi, and Chang, 2010]. If we assume that the duration of the time-window is $T_W = 80$ ms, T_W contains 16 and 80 time-slots for Mobile WiMAX and LTE respectively.

Most of the simulation results in this dissertation are based on 10 users having Rayleigh fading channels, whose average CNRs are distributed with an average CNR of 15 dB. The average CNR of each user (for most of the simulations in this thesis) is given in Table 2.1.

When the CNR of a user is assumed to be correlated from time-slot to time-slot, one user can be allocated many consecutive time-slots. It is therefore more difficult to fulfill throughput guarantees for all the users in a system that has strongly temporally correlated channels. For the simula-

user	average CNR	
i	$\bar{\gamma}_i$ [dB]	$\bar{\gamma}_i$
1	5.0000	3.1623
2	9.7712	9.4868
3	11.9897	15.8114
4	13.4510	22.1359
5	14.5424	28.4605
6	15.4139	34.7851
7	16.1394	41.1096
8	16.7609	47.4342
9	17.3045	53.7587
10	17.7875	60.0833

TABLE 2.1: Average CNRs of the 10 Rayleigh-distributed users distributed with an average of 15 dB.

tions in this section, we assume Jakes' correlation model given in Section 1.3.2.

Each value in the simulated TGVP curves is an average over 1000 and 500 Monte Carlo simulations, for the Mobile WiMAX and LTE systems respectively.

2.5 Numerical results

Figures 2.1 and 2.2 give a comparison of the approximate TGVP expression for the MCS and the NCS algorithms with the corresponding Monte Carlo simulated TGVPs for Mobile WiMAX and LTE based networks respectively, for 10 users with average CNRs given in Table 2.1. For simplification, we have assumed identical throughput guarantees, i.e. $B_i = B$ for all $i = 1, \cdots, 10$. The results from the approximate TGVP expression are based on the assumption that the time-slots are uncorrelated, while the Monte Carlo simulations are for users that have a correlated CNR from time-slot to time-slot.

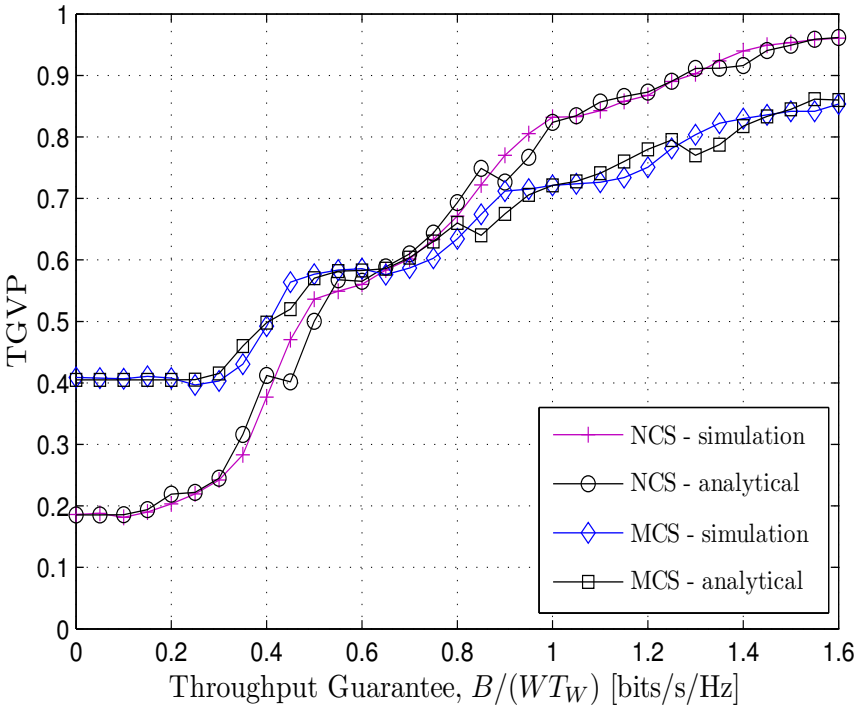


FIGURE 2.1: Comparison of the approximate TGVP expression with the Monte Carlo simulated TGVP for a network with 10 users. The curves are plotted for the Mobile WiMAX network with time-slot length T_{TS} = 5 ms. A time-window of length T_W = 80 ms corresponds to $K = 16$ time-slots.

The NCS algorithm (as expected) outperforms the MCS algorithm in terms of the TGVP performance. As shown in Figure 2.2, unlike the MCS, the NCS algorithm is able to provide unity throughput guarantee to all the users for a certain range. This is because it is a fairer algorithm, and thus gives more time-slots to the users with bad channels.

The tightness of the approximate TGVP expression is influenced by the number of time-slots K, by the length of a time-slot T_{TS}, and by the maximum Doppler frequency shift f_D or the user speed v for a constant carrier frequency f_c. In order to obtain a tight approximation, the TGVP should be calculated for a relatively large number of time-slots. For a fixed time-window T_W and a fixed $f_D = 100$ Hz (corresponding to $f_c = 1$ GHz and

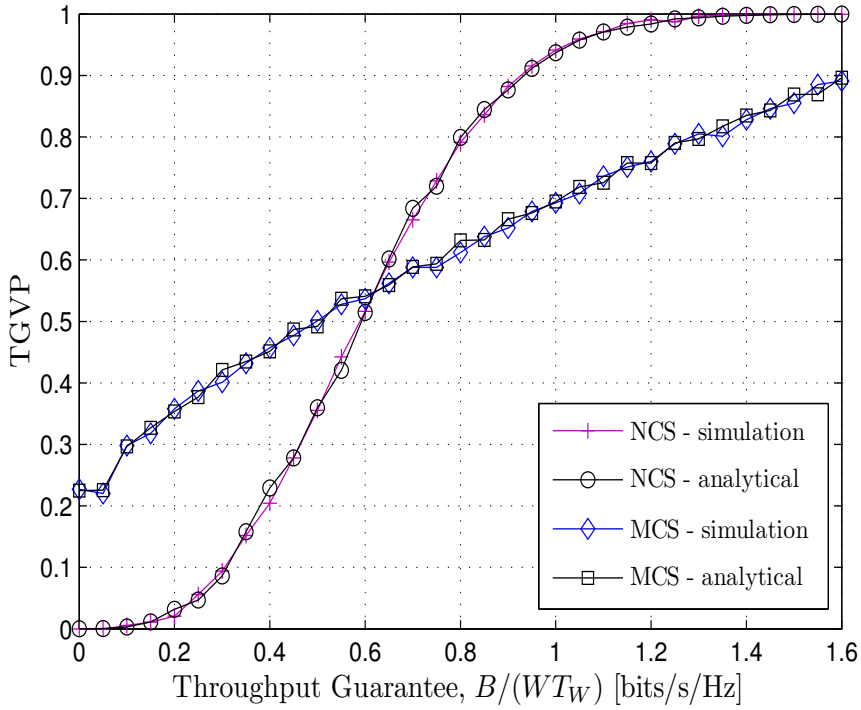


FIGURE 2.2: Comparison of the approximate TGVP expression with the Monte Carlo simulated TGVP for a network with 10 users. The curves are plotted for the LTE network with time-slot length $T_{\text{TS}} = 1$ ms. A time-window of length $T_W = 80$ ms corresponds to $K = 80$ time-slots.

$v = 30$ m/s), a larger K would mean a shorter time-slot duration, and therefore a higher correlation between the time-slots. In our work, we have calculated the analytical approximation of the TGVP by assuming uncorrelated time-slots. Therefore, we expect a less tight approximation for short time-slots. We see that the TGVP approximation for LTE network ($K = 80$ time-slots) is better than Mobile WiMAX ($K = 16$ time-slots). We can therefore conclude that the number of time-slots K within the time-window T_W will have a greater effect on the tightness of the approximate TGVP expression than the correlation between the shorter time-slots.

Next we observe the tightness of the approximate TGVP expression when we have a longer time-window, again for $f_D = 100$ Hz. We consider

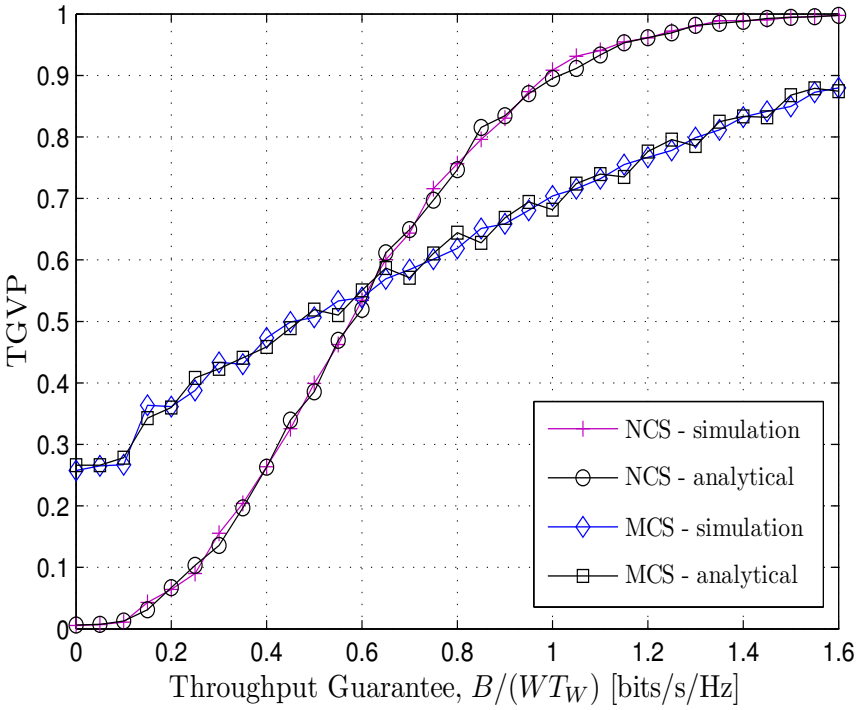


FIGURE 2.3: Comparison of the approximate TGVP expression with the Monte Carlo simulated TGVP for a network with 10 users. The curves are plotted for the Mobile WiMAX network with time-slot length T_{TS} = 5 ms. A time-window of length T_W = 240 ms corresponds to $K = 48$ time-slots.

Mobile WiMAX network consisting of 10 users with average CNRs given in Table 2.1. Compare Figures 2.1 and 2.3 where the numbers of time-slots within the time-window are 16 and 48 respectively. We see that a longer time-window has lead to a tighter TGVP approximation. This is because a longer time-window T_W has a higher number of time-slots K, and as mentioned before the tightness of the TGVP expression requires a relatively large number of time-slots. Note also that the overall TGVP performance of both the schemes is better for the second case (longer time-window). There are more time-slots now, thus making it possible for the MCS and the NCS algorithms to fulfill the throughput guarantees of more users.

Finally, we observe the accuracy of the approximate TGVP expression

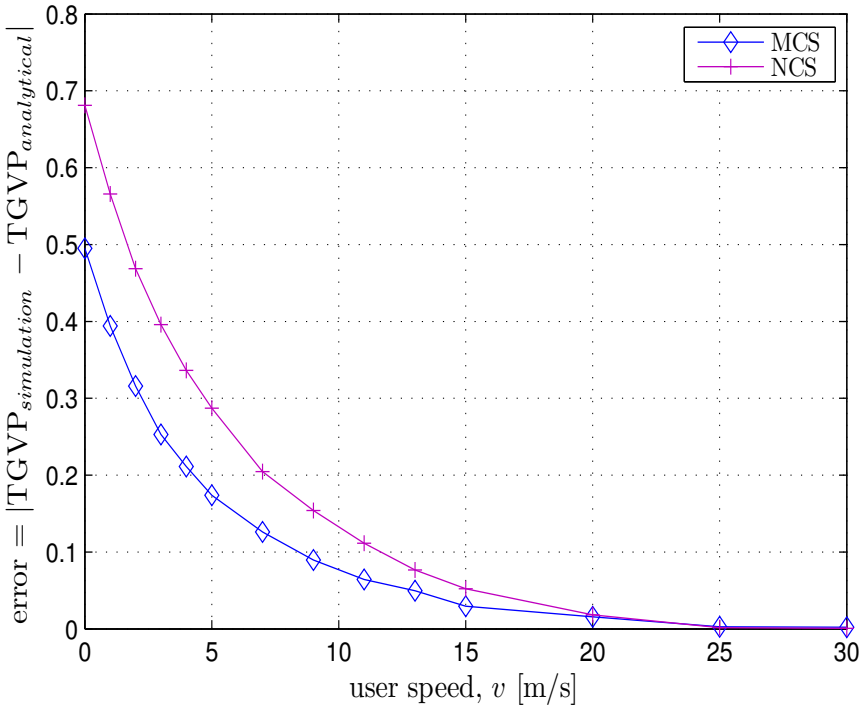


FIGURE 2.4: Effect of users' speed on the approximate TGVP expression for 10 users in a Mobile WiMAX network with identical throughput guarantees of $B/(WT_W) = 0.2$ bits/s/Hz, where $T_W = 80$ ms, corresponding to $K = 16$ time-slots.

for various user speeds. We again consider Mobile WiMAX network consisting of 10 users with average CNRs given in Table 2.1, and identical throughput guarantees of $B/(WT_W) = 0.2$ bits/s/Hz, where $T_W = 80$ ms. Figure 2.4 shows the difference (error) between the Monte Carlo simulated TGVP and the approximate TGVP expression for various users' speeds. We observe high errors at lower users' speeds. For fixed time-slot length, the temporal correlation of the channel is both dependent on the speed v of the users and on the carrier frequency f_c of the channel. For fixed f_c, lower user speed means lower Doppler frequency shift f_D, which results in a less rapidly changing user channel, and subsequently higher correlation between the time-slots. Therefore the TGVP of the network should be higher for lower users' speeds. However, we have calculated the analyti-

cal approximation of the TGVP by assuming uncorrelated time-slots. We thus conclude that the TGVP expression (when used for the MCS and NCS algorithm) is less accurate at lower speeds.

2.6 Summary

In this chapter, we have derived an approximate expression for the TGVP which can be obtained in a time-slotted wireless network with any scheduling policy with a given set of system parameters, known cumulants of the bits transmitted to or from the scheduled user in a time-slot, and a given distribution of the number of time-slots allocated to a user within a time-window. The violation of the throughput guarantee is treated as a rare event, and then Cramér's theorem is applied to derive the approximate TGVP expression. The simulations show that the number of time-slots has a greater effect on the tightness of the approximate TGVP expression than the correlated time-slots. We also observe that the approximate TGVP expression is tighter for longer time-windows. Furthermore, we note that the TGVP expression is more accurate for fast moving users.

Chapter 3

Scheduling Algorithms for Improved Throughput Guarantees

As already mentioned, a throughput guarantee is a well suited measure to quantify the QoS experienced by a user exactly. Both the network operators and the customers would like to have a scheduling algorithm that would promise as high a throughput guarantee as possible. Borst and Whiting have proved that a certain scheduling policy provides the highest throughput guarantee for wireless networks [Borst and Whiting, 2003]. However, they briefly argue that the rate distributions of the users are unknown and they have therefore not shown how this optimal scheduling policy can be found for users with differently distributed CNRs. In our work, we argue that for many scenarios the CNR distributions of the users can in fact be estimated, and that we hence can use these distributions to develop efficient scheduling algorithms for providing short-term throughput guarantees. In modern cellular standards like HSDPA, LTE and Mobile WiMAX [WiMAX Forum, 2006], much of the information needed for obtaining precise probability distribution estimates is already available. For example, in order to conduct adaptive coding and modulation, modern cellular networks have precise, real-time user CNR estimates. These channel quality estimates can also be used to obtain estimates of the probability distributions of the CNRs of each one of the users.

In this chapter, we formulate an optimization problem aimed at find-

ing an optimal scheduling algorithm that obtains maximum throughput guarantees in a wireless network. By building on the results in [Borst and Whiting, 2003] and by assuming that the distributions of the users' CNRs are known, we show how the solution to this optimization problem can be obtained numerically both when the throughput guarantees are (a) the same and (b) different for all the mobile users. We also propose two adaptive algorithms that improve the performance of the optimal algorithm for short time-windows. In real systems, some of the users are static users, while others are pedestrian or vehicular users. We therefore also analyze the performance of these algorithms for different time-slot correlations corresponding to different users' speeds. Furthermore, we also analyze the spectral efficiency and fairness of the proposed optimal scheduling algorithm in this chapter.

The rest of this chapter is organized as follows. In Section 3.1 we present the system model, and in Section 3.2 we formulate the optimization problem for obtaining the maximum throughput guarantee over a time-window. In Section 3.3 we show how to obtain the solution to this problem when all the users have identical throughput guarantees. The corresponding solution for heterogeneous throughput guarantees is discussed in Section 3.4. We describe the novel adaptive scheduling algorithms in Section 3.5. In Section 3.6, we derive the expressions for the parameters of the TGVP expression for the optimal scheduling algorithm. Spectral efficiency and fairness of this algorithm is considered in Section 3.7. In Section 3.8 we discuss some practical considerations, before presenting our numerical results in Section 3.9. A brief summary/discussion is given in Section 3.10.

This chapter is primarily based on [Rasool et al., 2011; Rasool and Øien, 2010a,b].

3.1 System Model

We consider a single base station that serves N users using TDM. The analysis conducted in this chapter is valid both for the uplink and the downlink; in either case we assume that the total available bandwidth for the users is W [Hz] and that the users have constant transmit power. Each user is assumed to estimate his own CNR perfectly, and before performing downlink scheduling the base station is assumed to receive these measurements from all the users. The base station also performs uplink scheduling based

on perfect channel estimates, and for each time-slot, the base station takes a scheduling decision and distributes this decision to the selected user before uplink transmission starts.

The communication channel between the base station and all the users is modeled by a flat, block-fading channel subject to AWGN noise; and it is further assumed that the communication channels corresponding to the different users fade independently. The block fade duration is assumed to equal one time-slot, and is denoted by T_{TS} [seconds]. We also assume that the CNR values corresponding to different time-slots are correlated. The correlation model used in our simulations is described in Section 1.3.2.

The average CNR of user i is denoted by $\bar{\gamma}_i$, and it is assumed to be constant for the time-window over which the throughput guarantees are calculated[1]. Without loss of generality, we assume that the user indices are assigned in a manner such that user 1 has the lowest average CNR, user 2 has the second lowest average CNR, and so on, down to user N, which has the highest average CNR.

We also assume that the probability distributions of the CNRs of each of the users are perfectly known (however, a known *joint* CNR distribution is *not* required). Another important assumption is that the population of backlogged users is constant and equal to N. It is also assumed that users always have data to send. We also assume that only one user can be scheduled in a time-slot.

3.2 The Optimization Problem

We now formulate an optimization problem aimed at obtaining the maximal throughput guarantee which can be promised to all the users within a time-window of T_W [seconds][2]. In this section, we assume that all the users are promised identical throughput guarantees, i.e. $B_i = B$ [bits] for all $i = 1, ..., N$. We have

$$T_i \bar{R}_i = B, \tag{3.1}$$

where T_i [seconds] is the total time allocated to user i over the time-window and $\bar{R}_i$ [bits/s] is the average rate for user i when he is transmitting or receiving.

[1] The justification of most of the assumptions in this section is given in Section 2.1.

[2] A similar optimization problem has also been formulated in [Borst and Whiting, 2003], and explored in [Hassel et al., 2007; Rasool et al., 2011; Rasool and Øien, 2010b].

Due to the TDM assumption, the sum of the T_is satisfies

$$\sum_{i=1}^{N} T_i = T_W. \tag{3.2}$$

From (3.1) and (3.2), we obtain

$$B = \frac{T_W}{\sum_{i=1}^{N} \frac{1}{\bar{R}_i}}. \tag{3.3}$$

Under the assumption that each T_i is long enough to make the time-window T_W appear infinitely long, (3.1) can also be written as

$$p(i) T_W \bar{R}_i = B, \tag{3.4}$$

where $p(i)$ is the *access probability* for user i within the time-window T_W. From (3.4) and (3.3), we get

$$p(i) = \frac{1}{\bar{R}_i \sum_{j=1}^{N} \frac{1}{\bar{R}_j}}. \tag{3.5}$$

The average rate $\bar{R}_i$ for user i when he is transmitting or receiving, can be written as

$$\bar{R}_i = W \int_0^\infty \log_2(1+\gamma) f_{\gamma *}(\gamma|i) d\gamma, \tag{3.6}$$

where $f_{\gamma *}(\gamma|i)$ is the PDF of the CNR of user i when this user is scheduled. Here, we have assumed that T_W is long enough and contains enough time-slots for the channel to reveal its ergodic properties, and that the Shannon capacity can be achieved.

From the equations above, the goal is to find a *scheduling algorithm* that gives the maximum throughput guarantee B that can be promised to all the users over the time-window T_W. Thus, (3.3) has to be maximized subject to the constraints given in (3.6), for $i = 1, \ldots, N$. In the next section, we show how to obtain this optimal scheduling algorithm.

3.3 Solution to the Optimization Problem

It was shown in [Borst and Whiting, 2003] that the following scheduling algorithm gives the solution to the optimization problem described in the previous section:

$$i^*(n) = \underset{1 \le i \le N}{\operatorname{argmax}} \left(\frac{r_i(n)}{\alpha_i} \right), \tag{3.7}$$

where $i^*(n)$ is the index of the user that is to be scheduled in time-slot n, $r_i(n)$ is the instantaneous rate of user i in time-slot n, and α_i is a constant to be optimized. This algorithm will be called *maximum throughput guarantee scheduling* (MTGS) algorithm in this dissertation.

In [Borst and Whiting, 2003], Borst and Whiting have not shown how the optimal α_is can be found. The method to find these α_is is suggested in [Hassel et al., 2007], under the assumptions that the PDFs of the users' channel gains are known, and that we have an ideal link adaptation protocol and block-fading. To obtain this solution, we define the random variable $S_i \stackrel{\triangle}{=} \frac{R_i}{\alpha_i}$, where R_i is the random variable describing the rate of user i. S_i is thus the scheduling metric of the algorithm, i.e. the metric that decides which user is going to be scheduled. For flat, block-fading AWGN channels, the maximal value of S_i within a time-slot (block) with CNR γ can be expressed as

$$S_i(\gamma) = \frac{W \log_2(1+\gamma)}{\alpha_i}. \tag{3.8}$$

In real-life systems we can come close to this maximum value of S_i by utilizing efficient link adaptation and (close-to-)capacity-achieving codes [Hassel et al., 2007]. The PDF for the normalized rate $S_i = s$ for user i can be written as

$$f_{S_i}(s) = \left. \frac{f_{\gamma_i}(\gamma)}{\frac{dS_i(\gamma)}{d\gamma}} \right|_{\gamma = 2^{\frac{s \cdot \alpha_i}{W}} - 1}, \tag{3.9}$$

where $f_{\gamma_i}(\gamma)$ is the PDF of the CNR of user i. Assuming Rayleigh faded channel gains for all the users[3], the PDF $f_{\gamma_i}(\gamma)$ is then given in (1.4). We have

$$f_{S_i}(s) = \frac{\alpha_i \ln(2)}{W \bar{\gamma}_i} 2^{\frac{s \cdot \alpha_i}{W}} e^{-\frac{2^{\frac{s \cdot \alpha_i}{W}} - 1}{\bar{\gamma}_i}}. \tag{3.10}$$

The corresponding CDF for the normalized rate $S_i = s$ can be expressed as

$$\begin{aligned} F_{S_i}(s) &= \int_0^s f_{S_i}(x) dx \\ &= 1 - e^{-\frac{2^{\frac{s \cdot \alpha_i}{W}} - 1}{\bar{\gamma}_i}}. \end{aligned} \tag{3.11}$$

[3] In principle we could have used different CNR distributions (e.g. Rayleigh, Rice, Nakagami) for the different users. However, for simplicity reasons, we have assumed that all the users have Rayleigh distributed channel gains.

The access probability of user i can now be expressed as [Yang and Alouini, 2006]

$$p(i) = \int_0^\infty f_{S_i}(s) \prod_{\substack{j=1 \\ j \neq i}}^{N} F_{S_j}(s)\, ds. \tag{3.12}$$

Using *Bayes' rule*, the PDF of S_i when user i is scheduled is given as

$$f_{S_i}(s|i) = \frac{f_{S_i}(s)}{p(i)} \prod_{\substack{j=1 \\ j \neq i}}^{N} F_{S_j}(s). \tag{3.13}$$

We can also express the expected value of S_i conditioned on user i being scheduled, as

$$\begin{aligned} \mathbb{E}[S_i|i] = \frac{\mathbb{E}[R_i|i]}{\alpha_i} &= \frac{\bar{R}_i}{\alpha_i} \\ &= \int_0^\infty s\, f_{S_i}(s|i)\, ds. \end{aligned} \tag{3.14}$$

Combining (3.5), (3.12), and (3.14) we obtain $3N$ equations in $3N$ unknowns, and can thus find the values for the $p(i)$s, the $\bar{R}_i$s, and the α_is [Hassel et al., 2007]. A solution can be obtained by using numerical integration together with an algorithm for solving sets of nonlinear equations. This can for example be achieved in MATLAB by using the functions `quad` and `fsolve`.

Since the MTGS algorithm maximizes B, it is expected to yield higher values of B than any of the other scheduling algorithms. However, it should be remembered that it is implicitly assumed in (3.1) that the average rate of the users over the time-window equals their expected throughput. This will only be true for infinitely long time-window T_W, containing infinitely many time-slots. The solution is thus suboptimal for short time-windows. In Section 3.5 we therefore propose two adaptive scheduling algorithms that improve the short-term performance.

Note that when $\alpha_i = \alpha$ for all $i = 1, \ldots, N$, the MTGS algorithm given in (3.7) reduces to the MCS algorithm, which schedules the user with the highest CNR, and hence the highest rate.

3.4 Optimization for Heterogeneous Throughput Guarantees

When the throughput guarantees are different from user to user, we can again use the scheduling policy corresponding to (3.7), but with a different set of α_is to obtain the optimal bit allocation. By using B_i [bits] to denote the throughput guarantee for user i during the time-window T_W, we obtain

$$T_i \bar{R}_i = B_i. \tag{3.15}$$

(3.2) becomes

$$\sum_{i=1}^{N} \frac{B_i}{\bar{R}_i} = T_W. \tag{3.16}$$

For a finite but long time-window T_W, we have[4]

$$p(i) T_W \bar{R}_i \approx B_i. \tag{3.17}$$

From (3.16) and (3.17), we obtain the following expression for $p(i)$:

$$p(i) \approx \frac{B_i}{\bar{R}_i \sum_{j=1}^{N} \frac{B_j}{\bar{R}_j}} \tag{3.18}$$

We can now fix the throughput guarantees B_i of up to $N-1$ users and maximize the remaining throughput guarantees by solving the set of $3N$ equations resulting from (3.12), (3.14) and (3.18). To be able to solve this optimization problem, we can for example additionally constrain the remaining users (those with non-fixed B_is) to have equal throughput guarantees. It is also important to note that setting fixed throughput guarantees that are too ambitious will yield an optimization problem with no solution – meaning that such throughput guarantees are not achievable by the system. Of course it only makes sense to set fixed throughput guarantees that are achievable by the system [Hassel et al., 2007].

3.5 Improving the Short-Term Performance

As already mentioned, the MTGS algorithm is only efficient when the throughput guarantees are promised over a long time-window T_W containing many

[4]For an infinitely long time-window T_W and an ergodic channel, (3.17) and (3.18) become equalities.

time-slots. To fulfill throughput guarantees for shorter time-windows with fewer time-slots, we propose two adaptive scheduling schemes in this section.

3.5.1 Adaptive Maximum Throughput Guarantee Scheduling Algorithm 1 (AMTGS1)

The parameters α_i, obtained in the previous sections, are optimized such that the throughput guarantees should be fulfilled independently of the time instants at which T_W starts or ends. In this subsection we develop another algorithm that only aims at fulfilling the throughput guarantees within the duration of a *fixed* time-window T_W.

To improve performance for shorter time-windows, we need to *adapt* the values of the parameters α_i to the actual resource allocation that has already been done within the finite time-window T_W. This adaptation can optimally be done during each time-slot by using the approach of the previous section with B_i/T_W replaced by

$$\frac{B_i'}{T_W'} = \frac{B_i - B_{in}}{T_W - T_n},$$

where B_{in} is the number of bits assigned to user i after n time-slots within the time-window T_W, and $T_n = nT_{\mathrm{TS}}$ [Hassel et al., 2007]. The adaptation of the parameters α_i should in many cases be performed in time intervals of less than a millisecond. Since it can be difficult to conduct the optimal optimization described above in such a short time, we here propose the AMTGS1 algorithm as an alternative[5]:

$$i^*(n) = \underset{1\le i\le N}{\operatorname{argmax}} \left(\vartheta_i(n-1)\frac{r_i(n)}{\alpha_i} \right), \tag{3.19}$$

where $\vartheta_i(n)$ is the ratio

$$\vartheta_i(n) = \frac{\max(0, B_i - B_{in})}{T_W - T_n}\frac{T_W}{B_i}. \tag{3.20}$$

The rationale behind this scheduling algorithm is as follows: The value of $\vartheta_i(n)$ expresses the normalized share of the throughput guarantee that is to be fulfilled in the remaining $K - n$ time-slots of the time-window T_W.

[5]Originally proposed in [Hassel et al., 2007].

- If the rate guarantee is already fulfilled, the value of $\vartheta_i(n)$ is zero, which means that the user in question is not selected in the remaining $K - n$ time-slots.

- If a user has been allocated exactly $\frac{B_i T_n}{T_W}$ bits after n time-slots, the value of $\vartheta_i(n)$ will be unity, which means that this user will be scheduled with the same weights as for the MTGS algorithm.

- For the case where the number of allocated bits after n time-slots is lower than $\frac{B_i T_n}{T_W}$ bits, the value of $\vartheta_i(n)$ will be above unity, which means that the user is given higher priority compared to the MTGS algorithm.

- Likewise, a user is given lower priority if he has been allocated more than $\frac{B_i T_n}{T_W}$ bits after n time-slots.

The priority is thus determined by the urgency of fulfilling the throughput guarantee within the remainder of the time-window.

A similar strategy has also been employed in [Kastrinogiannis and Papavassiliou, 2008] for improving short-term throughput of utility-based scheduling in CDMA wireless networks.

3.5.2 Adaptive Maximum Throughput Guarantee Scheduling Algorithm 2 (AMTGS2)

The problem with the AMTGS1 algorithm is that it can only fulfill the throughput guarantees when the placement of the window is fixed. That is, for every new time-window, the algorithm starts over again and tries to achieve the throughput guarantees. This means that the throughput guarantees cannot be promised within time-windows with a different duration or a different placement than that used by the algorithm. The consequence of this approach is that we may have to adjust the time-window T_W to the bit-streams from different speech and video codecs.

In this subsection, we propose another adaptive scheduling algorithm, i.e. AMTGS2 [6], that overcomes the problem of fixed window placement of AMTGS1. Furthermore, the scheduling algorithm is also simpler in implementation. The AMTGS2 algorithm works as follows:

[6] Initially proposed in [Rasool and Øien, 2010b].

For promised throughput guarantees B_i, a user $i^*(n)$ is selected that has a maximum

$$i^*(n) = \underset{1 \le i \le N}{\operatorname{argmax}} \left(v_i(n-1) \frac{r_i(n)}{\alpha_i} \right), \tag{3.21}$$

where $v_i(n)$ is given as

$$v_i(n) = \begin{cases} 0 & \text{if } B_{in} \ge B_i \\ 1 & \text{otherwise} \end{cases} \tag{3.22}$$

Here, B_{in} is the total number of bits assigned to user i after n time-slots. The rationale behind this scheduling algorithm is very simple, and is given as follows:

- If the throughput guarantee of user i is already fulfilled, it is not selected in the remaining time-slots, i.e., the value of $v_i(n)$ is set to zero.
- For all the other users, $v_i(n) = 1$ so that among them, a user j is selected with maximum $r_j(n)/\alpha_j$.

Note that the AMTGS2 algorithm is independent of the duration and placement of the time-window T_W.

We can intuitively say that the offline parameter α_i increases the throughput fairness of the system, whereas the online parameters ϑ_i and v_i improve the corresponding short-term performance of the system.

3.6 The TGVP Expression's Parameters for the MTGS Algorithm

The number of time-slots allocated to user i within a time-window of K time-slots is distributed according to the binomial distribution expressed in (2.14) with $p(i)$ given in (3.12).

The first moment m_1 (the mean value for $b_{i,j}$ where $b_{i,j}$ is the number of bits transmitted to/from user i in jth time-slot he is scheduled) for the MTGS algorithm is derived as follows:

$$\begin{aligned} m_1 &= \mathbb{E}[b_{i,j}] \\ &= WT_{\mathrm{TS}} \int_0^\infty \log_2(1+\gamma) f_{\gamma^*}(\gamma|i) d\gamma \\ &= \alpha_i T_{\mathrm{TS}} \int_0^\infty s\, f_{S_i}(s|i)\, ds. \end{aligned} \tag{3.23}$$

Using (3.14),

$$\mathbb{E}[b_{i,j}] = T_{\mathrm{TS}}\bar{R}_i. \tag{3.24}$$

Similarly, the second moment m_2 of the number of bits $b_{i,j}$ transmitted to or from user i can be obtained as follows:

$$\begin{aligned} m_2 &= \mathbb{E}[b_{i,j}^2] \\ &= (WT_{\mathrm{TS}})^2 \int_0^\infty (\log_2(1+\gamma))^2 f_{\gamma*}(\gamma|i) d\gamma \\ &= (\alpha_i T_{\mathrm{TS}})^2 \int_0^\infty s^2 \, f_{S_i}(s|i) \, ds. \end{aligned} \tag{3.25}$$

We analyze the correctness of these derived parameters of the TGVP expression in Appendix C.

3.7 Spectral Efficiency and Fairness of the MTGS Algorithm

The spectral efficiency can be analyzed with the help of ASSE which was defined in Section 1.8.2. Using (3.6) and (3.14), the ASSE of the MTGS algorithm can be expressed as:

$$\begin{aligned} \mathrm{ASSE}_{MTGS} &= \sum_{i=1}^{N} p(i) \int_0^\infty \log_2(1+\gamma) f_{\gamma*}(\gamma|i) \, d\gamma \\ &= \frac{1}{W} \sum_{i=1}^{N} p(i)\alpha_i \int_0^\infty s \, f_{S_i}(s|i) \, ds. \end{aligned} \tag{3.26}$$

To analyze the fairness of the MTGS algorithm, we make use of the *Jain's Fairness Index* (JFI) defined in Section 1.8.3. Choosing X in (1.20) to be the number of time-slots allocated to the users, we can define the *time-slot fairness* [Hassel et al., 2006]:

$$\mathrm{JFI}_{\mathrm{TS}} = \frac{(\mathbb{E}_K[X_{\mathrm{TS}}])^2}{\mathbb{E}_K[X_{\mathrm{TS}}^2]}, \tag{3.27}$$

where X_{TS} is a random variable that describes the number of time-slots allocated to an arbitrary user within a window of K time-slots. This random variable is discrete and can take on the values $k = 1, ..., K$. The first moment of the time-slot allocation is given as:

$$\mathbb{E}_K[X_{\mathrm{TS}}] = \frac{1}{N} \sum_{i=1}^{N} \sum_{k=1}^{K} k \, p_K(k|i). \tag{3.28}$$

Similarly, we can find the second moment of the time-slot allocation to be:

$$\mathbb{E}_K[X_{\mathrm{TS}}^2] = \frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K} k^2 p_K(k|i). \tag{3.29}$$

where $p_K(k|i)$ is given in (2.14) with $p(i)$ given in (3.12).

Another useful criterion for fairness can be *throughput fairness*, which can be defined as [Hassel et al., 2006]:

$$\mathrm{JFI}_{\mathrm{TP}} = \frac{(\mathbb{E}_K[X_{\mathrm{TP}}])^2}{\mathbb{E}_K[X_{\mathrm{TP}}^2]}, \tag{3.30}$$

where X_{TP} [bits per time-window per Hz] is a random variable describing the throughput allocated to an arbitrary user within K time-slots. The first moment of the throughput allocation for the MTGS algorithm can be written as follows:

$$\begin{aligned}\mathbb{E}_K[X_{\mathrm{TP}}] &= \frac{1}{N}\sum_{i=1}^{N}\sum_{k=0}^{K} p_K(k|i)\int_0^{\infty} k\log_2(1+\gamma) f_{\gamma^*}(\gamma|i)d\gamma \\ &= \frac{1}{WN}\sum_{i=1}^{N}\alpha_i\sum_{k=0}^{K} k p_K(k|i)\int_0^{\infty} s\, f_{S_i}(s|i)\, ds. \end{aligned}\tag{3.31}$$

Similarly, the second moment of the throughput allocation can be obtained as:

$$\begin{aligned}\mathbb{E}_K[X_{\mathrm{TP}}^2] &= \frac{1}{N}\sum_{i=1}^{N}\sum_{k=0}^{K} p_K(k|i)\int_0^{\infty} (k\log_2(1+\gamma))^2 f_{\gamma^*}(\gamma|i)d\gamma \\ &= \frac{1}{W^2N}\sum_{i=1}^{N}\alpha_i^2\sum_{k=0}^{K} k^2 p_K(k|i)\int_0^{\infty} s^2\, f_{S_i}(s|i)\, ds. \end{aligned}\tag{3.32}$$

3.8 Practical Considerations and Simulation Parameters

In this section, we discuss some practical issues as well as realistic system parameters. Interested readers are referred to [Hassel et al., 2007] for a detailed discussion.

For the wireless standards HSDPA and Mobile WiMAX, the time-slot lengths for the downlink are 2 and 5 ms respectively [WiMAX Forum, 2006]. The corresponding time-slot length for the 3GPP LTE network is

1 ms [Freescale Semiconductor, 2007]. The European IST research project WINNER I has suggested a time-slot duration of 0.34 ms for a future wireless system [Sternad et al., 2006]. If we assume that $T_W = 80$ ms, the time-window contains 16, 40, 80 and 235 time-slots for Mobile WiMAX, HSDPA, LTE and WINNER I, respectively.

The results are shown for 10 users having Rayleigh fading channels whose average CNRs are distributed with an average of 15 dB. The average CNR of each user is given in Table 2.1.

It is more difficult to fulfill throughput guarantees for all the users in a system that has strongly temporally correlated channels, since one user can be allocated many consecutive time-slots. For the simulations in the next sections, we assume (modified) Jakes' correlation model given in Section 1.3.2.

If the average CNR of one or more users change or the CNR distribution of one or more users change, e.g., from Rayleigh to Rice, the whole optimization problem has to be solved again to obtain new values for the α_is, which is a feasible task. It should however be noted that the adaptive factors $\vartheta_i(n)$ and $v_i(n)$ are independent of the CNR distributions.

Each value in the plots is an average over 1000, 500, 500, and 250 Monte Carlo simulations for the Mobile WiMAX, HSDPA, LTE, and WINNER I systems respectively.

3.9 Numerical Results

3.9.1 Identical Throughput Guarantees

In this section, we consider the case where all the users are promised identical throughput guarantees B/T_W, where $T_W = 80$ ms. The optimized values of $p(i)$, $\bar{R}_i$ and α_i with identical throughput guarantees for 10 Rayleigh-distributed users are given in Table 3.1. From (3.4), we can see that

$$\frac{B_{opt}}{WT_W} = \frac{p(i)\bar{R}_i}{W},$$

for the optimized values of $p(i)$ and $\bar{R}_i$. It is easily seen by using the values in Table 3.1 to calculate the product $p(i)\bar{R}_i$ for $1 \leq i \leq 10$, that $B_{opt}/(WT_W) = 0.5675$ bits/s/Hz for all the users for this particular example.

i	$p(i)$	$\bar{R}_i$ [bit/s]	α_i
1	0.1804	3.1466	2.7519
2	0.1202	4.7194	4.5107
3	0.1042	5.4442	5.3549
4	0.0959	5.9176	5.9163
5	0.0905	6.2686	6.3380
6	0.0867	6.5488	6.6760
7	0.0837	6.7808	6.9581
8	0.0813	6.9789	7.2004
9	0.0794	7.1518	7.4127
10	0.0777	7.3048	7.6017

TABLE 3.1: The values of $p(i)$, $\bar{R}_i$ and α_i with identical throughput guarantees, for 10 Rayleigh-distributed users whose average CNRs are given in Table 2.1.

The correlation between the different time-slot CNRs is being described by the modified Jakes' model with $f_c = 1$ GHz and a user speed of $v = 30$ m/s. It should be noted that this correlation will be stronger for short time-slots than for long time-slots.

Figures 3.1-3.4 show the TGVP performance in networks that are respectively based on Mobile WiMAX, HSDPA, LTE and WINNER I. For these plots we have assumed that only one user can be scheduled in a time-slot. As mentioned earlier, we focus on the TGVP here since a throughput guarantee in most cases cannot be given with absolute certainty. Note that the TGVP performance of the algorithms close to TGVP$= 0$ (almost certainly guaranteed throughput performance) is the most interesting. We compare the new scheduling policies to five other algorithms, namely round robin scheduling (RR), maximum CNR scheduling (MCS), normalized CNR scheduling (NCS), proportional fair scheduling (PFS) and the *adaptive* algorithm proposed by Borst and Whiting in [Borst and Whiting, 2003]. The RR,

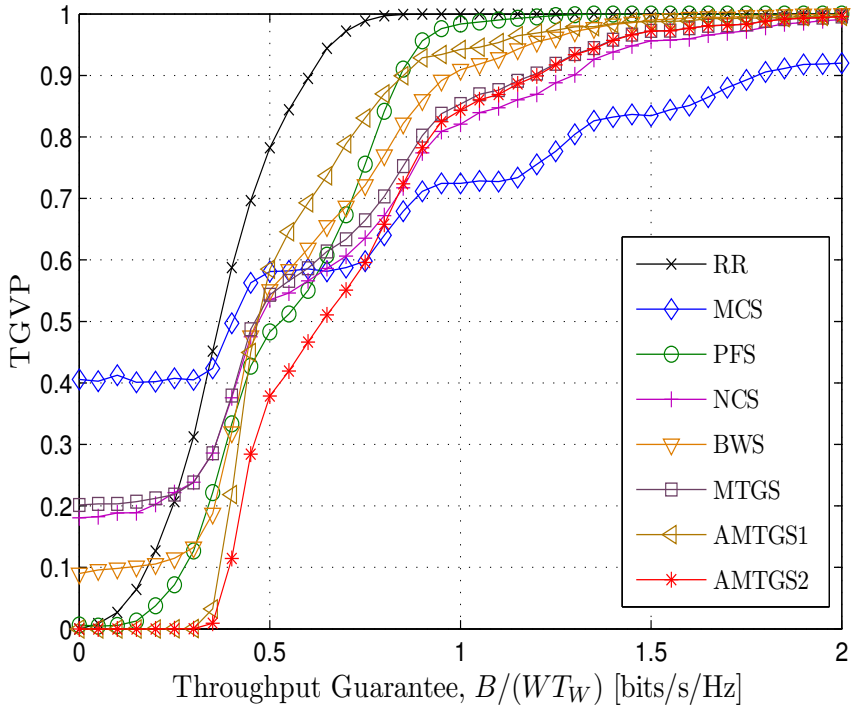


FIGURE 3.1: TGVP for 10 users in a Mobile WiMAX network with identical throughput guarantees. Plotted for a time-window $T_W = 80$ ms that contains 16 time-slots. Each value in the plot is an average over 1000 Monte Carlo simulations.

MCS, NCS and PFS algorithms are described in Sections 1.5.1, 1.6.1, 1.6.2 and 1.6.3 respectively. For our simulations, we have implemented the PFS algorithm as described in [Viswanath et al., 2002], with the time-constant $t_c = T_W$ and with the initial average rate for each user equal to the theoretical average rate for this user. The adaptive Borst and Whiting scheduling (BWS) algorithm is implemented as described in [Borst and Whiting, 2003, p. 575] with step size $\delta(n) = 0.5 * 0.9^n$, where n denotes the nth "reset". The "reset" is triggered when the throughputs of all the users become above-average. Borst and Whiting denote the process of updating the weights as "price updates". In our simulations, the "price updates" of this algorithm are done every 10th time-slot. In every update, the weights of the users are re-adjusted depending on whether their throughputs are above-average or

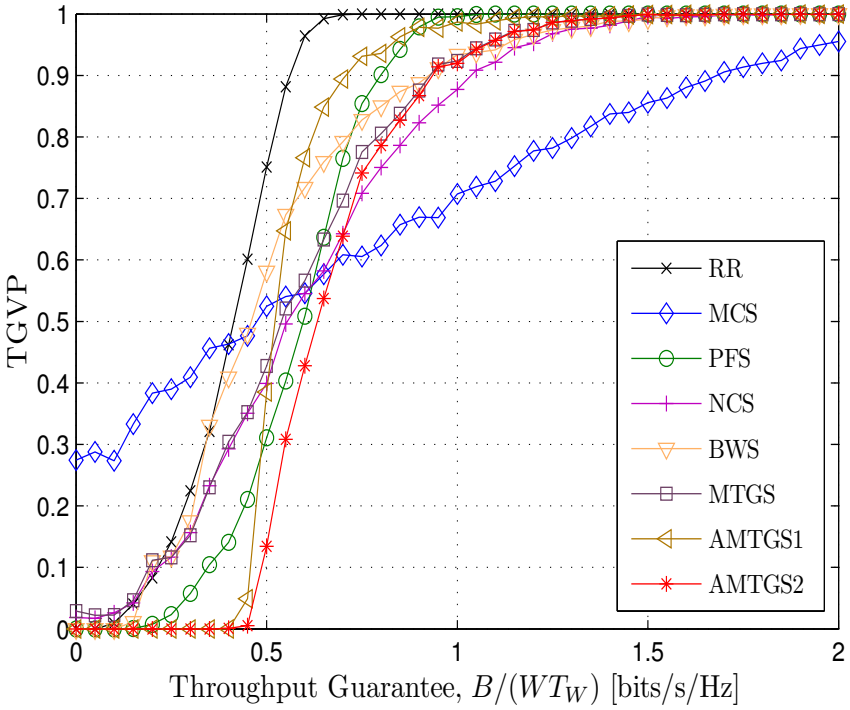


FIGURE 3.2: TGVP for 10 users in an HSDPA network with identical throughput guarantees. Plotted for a time-window $T_W = 80$ ms that contains 40 time-slots. Each value in the plot is an average over 500 Monte Carlo simulations.

below-average. To investigate the performance of the adaptive updating of the weights for this algorithm, we have used the optimal weights as initial weights.

Figures 3.1-3.4 show the TGVP as a function of $B/(WT_W)$ for a time-window of respectively 16, 40, 80 and 235 time-slots. We see that our adaptive algorithms (AMTGS1 and AMTGS2) perform better than all the other algorithms for all the cases. It should also be noted that since the WINNER I system has many time-slots within the time-window of 80 ms, the two adaptive algorithms obtain a throughput guarantee that is very close to the optimal throughput guarantee of 0.5675 bits/s/Hz for this system. It is also interesting to observe that the throughput guarantee that can be promised with close to unity probability with the adaptive algorithms is more than

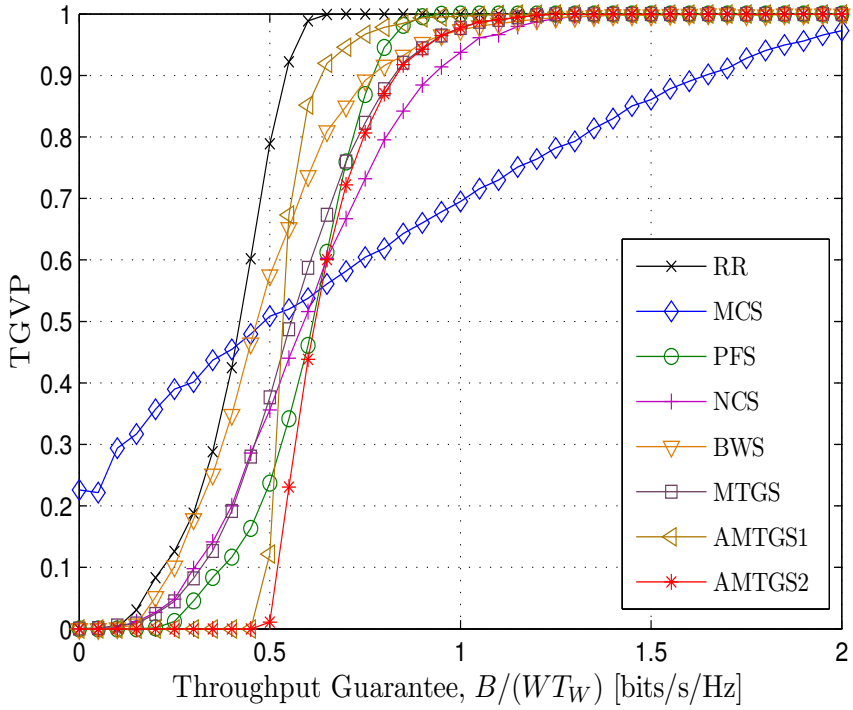


FIGURE 3.3: TGVP for 10 users in an LTE network with identical throughput guarantees. Plotted for a time-window $T_W = 80$ ms that contains 80 time-slots. Each value in the plot is an average over 500 Monte Carlo simulations.

twice as large as for the PFS algorithm for all the four systems.

It should be noted that our non-adaptive optimal algorithm (i.e. MTGS) also performs better than all the other well-known algorithms for the case where the time-window contains 235 time-slots (WINNER I). The reason for this is that the MTGS algorithm is designed for long time-windows containing many time-slots.

3.9.2 Heterogeneous Throughput Guarantees

Analyzing the TGVP for a network where the users have heterogeneous throughput guarantees requires a significant number of plots. We shall only consider an LTE based system, and to simplify the analysis, we fix the

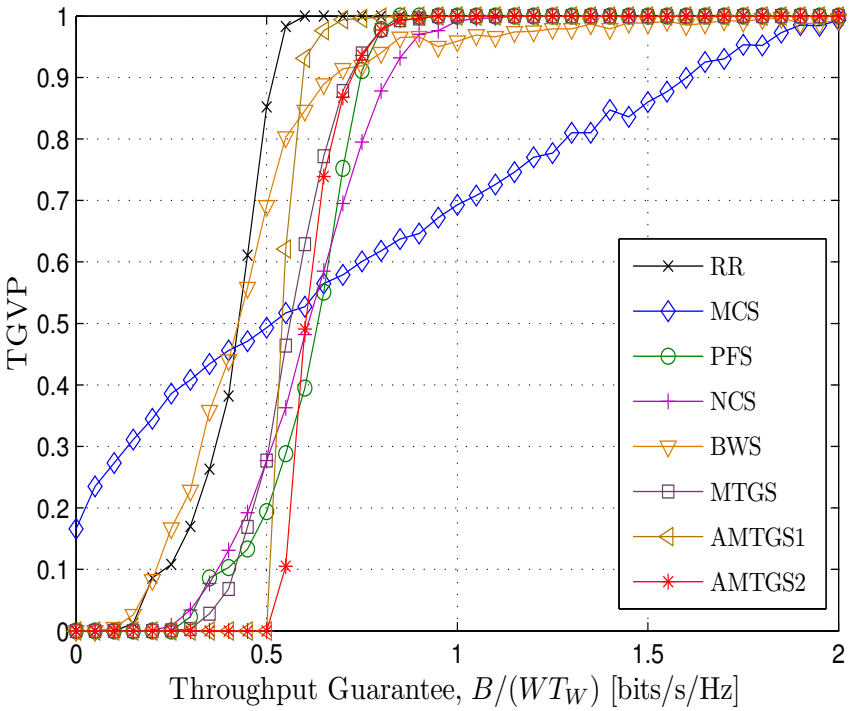


FIGURE 3.4: TGVP for 10 users in a WINNER I network with identical throughput guarantees. Plotted for a time-window $T_W = 80$ ms that contains 235 time-slots. Each value in the plot is an average over 250 Monte Carlo simulations.

throughput guarantees B_i/WT_W of 4 users to the same value, and maximize the throughput guarantees for the remaining users by solving the set of $3N$ equations resulting from (3.12), (3.14) and (3.18). We further constrain the users with non-fixed B_is to have equal throughput guarantees. The two sets of 4 users with fixed throughput guarantees, that we use in this section, are $\{1,2,3,4\}$ and $\{7,8,9,10\}$. These sets correspond to the users with low and high average CNRs respectively (refer to Table 2.1).

We first fix $B_i/WT_W = 0.3$ bits/s/Hz for the 4 users and try to maximize the throughput guarantee B/WT_W for the remaining six users. Note that these fixed throughput guarantees are lower than the $B_{opt}/WT_W = 0.5675$ bits/s/Hz of the case of identical throughput guarantees. From Figures 3.5 and 3.6, we observe that the two adaptive algorithms outperform all the

i	α_i for TG [bits/s/Hz]				
	all users	user $1-4$		user $7-10$	
	$B = 0.5675$	$B_i = 0.3$	$B_i = 0.7$	$B_i = 0.3$	$B_i = 0.7$
1	2.7519	3.4767	2.4735	2.6774	2.8541
2	4.5107	5.3770	4.1833	4.5083	4.6161
3	5.3549	6.2913	5.0068	5.3908	5.4615
4	5.9163	6.9008	5.5549	5.9786	6.0239
5	6.3380	6.5587	6.3572	6.4203	6.4462
6	6.6760	6.9223	6.6894	6.7745	6.7847
7	6.9581	7.2254	6.9670	7.7166	6.8229
8	7.2004	7.4855	7.2056	7.9750	7.0647
9	7.4127	7.7133	7.4147	8.2017	7.2766
10	7.6017	7.9159	7.6010	8.4037	7.4651

TABLE 3.2: Values of parameters α_is of the 10 Rayleigh-distributed users distributed with an average of 15 dB, for identical and heterogeneous throughput guarantees.

other algorithms. The performance of the MTGS algorithm is worse as compared to the scenario with identical throughput guarantees. Since this scheme is designed for a long time-window, it assumes that users that require 0.3 bits/s/Hz will get it in the long run. Therefore the selection of α_i's is such that lesser weight (higher value of α_i for these users compared to identical case) is given to these users, in order to maximize the throughput of the remaining users, as can be seen from Table 3.2. The performance of the adaptive algorithms is better than the previous case because 4 users require a throughput guarantee that is less than $B_{opt}/WT_W = 0.5675$ bits/s/Hz of the identical throughput guarantees case. The system allocates these extra bits to the remaining six users.

Next, we fix $B_i/WT_W = 0.7$ bits/s/Hz for the 4 users. This case is shown in Figures 3.7 and 3.8. Since we have selected fixed throughput guarantees higher than the $B_{opt}/WT_W = 0.5675$ bits/s/Hz of the identical throughput guarantees case, the performance of all the algorithms is

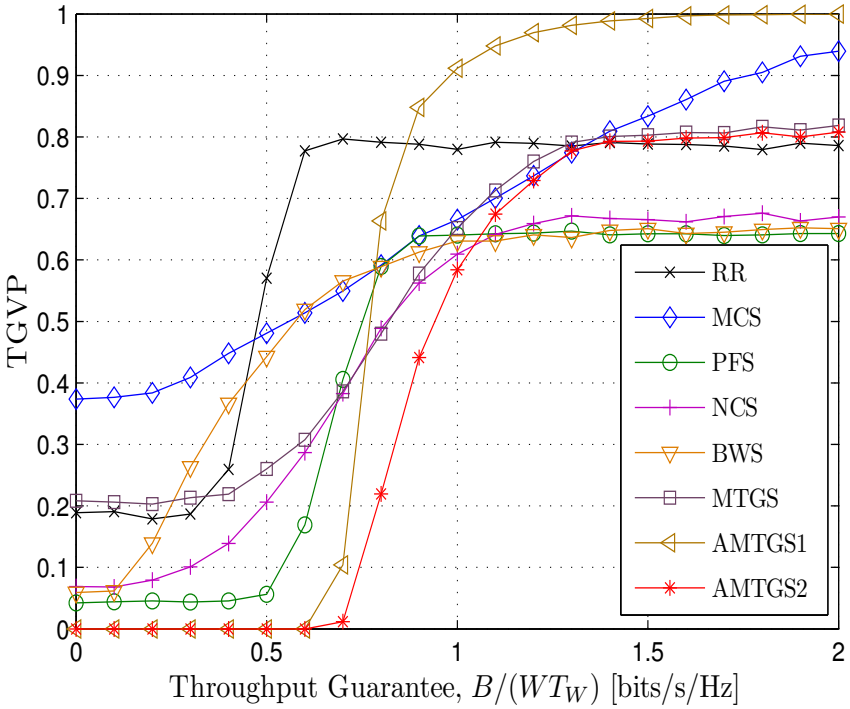


FIGURE 3.5: TGVP for 10 users in an LTE network. Throughput guarantees of users 1,2,3,4 are fixed to 0.3 bits/s/Hz and that of the remaining users is given by B/WT_W. Each value in the plot is an average over 500 Monte Carlo simulations.

bound to suffer. However, it is interesting to observe that our novel adaptive optimal scheduling algorithms are still able to provide close to unity throughput guarantee, i.e. TGVP $\cong$ 0 up to a certain maximum rate. In fact, they are the only algorithms that are able to do so in this case.

3.9.3 Effect of Temporal Correlation on TGVP

One user can be allocated many consecutive time-slots if the CNR of the users are correlated from time-slot to time-slot. It is therefore more difficult to fulfill throughput guarantees for all the users in a system that has strongly temporally correlated channels. As mentioned before, the temporal correlation of the channel is both dependent on the speed v of the

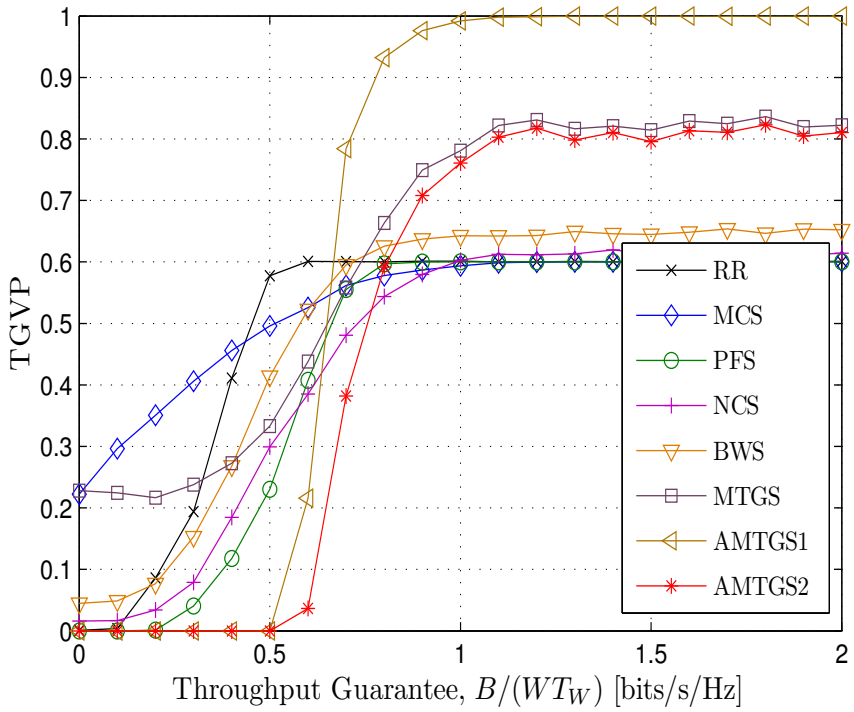


FIGURE 3.6: TGVP for 10 users in an LTE network. Throughput guarantees of users 7,8,9,10 are fixed to 0.3 bits/s/Hz and that of the remaining users is given by B/WT_W. Each value in the plot is an average over 500 Monte Carlo simulations.

users and on the carrier frequency f_c of the channel. For fixed f_c, higher user speed means higher Doppler frequency shift f_D, which results in a more rapidly changing user channel, and subsequently lower correlation between the time-slots. Thus we expect a better TGVP performance at higher speeds[7]. In this section, we observe (Figure 3.9) the effect of users' speed on the TGVP performance of different algorithms by considering a Mobile WiMAX network consisting of 10 users given in Table 2.1 and identical throughput guarantees of $B/(WT_W) = 0.2$ bits/s/Hz. We again use the modified Jakes' correlation model with $f_c = 1$ GHz.

As expected, the TGVP performance of the Round Robin scheduler re-

[7]This holds given our idealized assumption of perfect CSI - however this might be harder to obtain at high speeds.

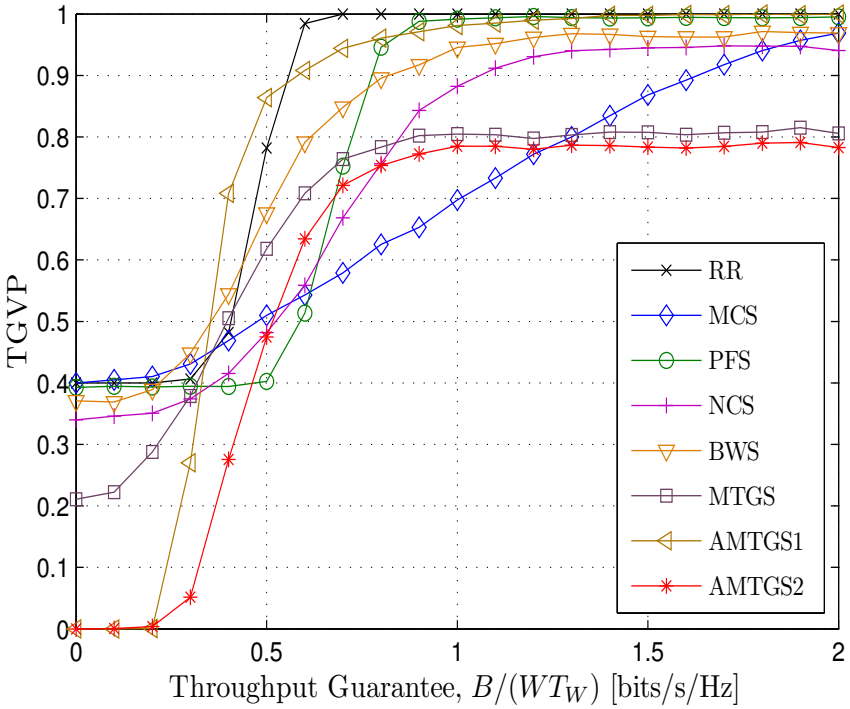


FIGURE 3.7: TGVP for 10 users in an LTE network. Throughput guarantees of users 1,2,3,4 are fixed to 0.7 bits/s/Hz and that of the remaining users is given by B/WT_W. Each value in the plot is an average over 500 Monte Carlo simulations.

mains the same for various users' speeds since the users are selected irrespective of their CNRs. The PFS algorithm does not suffer much at lower speeds since it takes into account the average throughput of all the users. The performance of the MCS and the MTGS algorithms at lower speeds deteriorate because they are not able to allocate enough time-slots to all the users due to strong temporal correlation. The performances of both the adaptive algorithms (AMTGS1 and AMTGS2) at lower users' speeds remain the best. This is due to the fact that these scheduling algorithms always ignore the users who have already received their share, and provide sufficient time-slots to other users to fulfill their throughput guarantees within the remaining time-window.

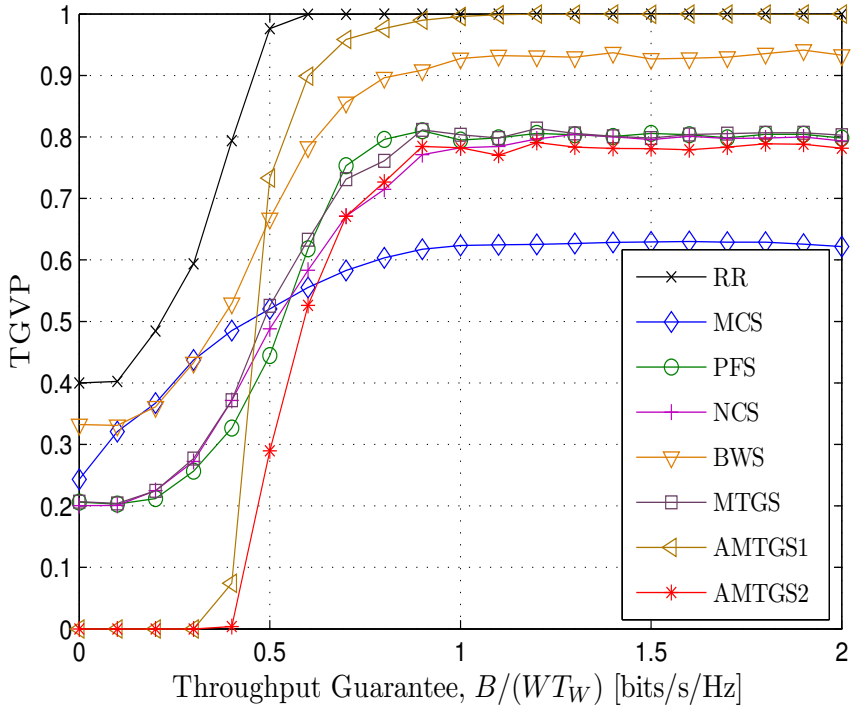


FIGURE 3.8: TGVP for 10 users in an LTE network. Throughput guarantees of users 7,8,9,10 are fixed to 0.7 bits/s/Hz and that of the remaining users is given by B/WT_W. Each value in the plot is an average over 500 Monte Carlo simulations.

3.9.4 ASSE Performance

In this section, we compare the ASSE performance of the MTGS algorithm with that of the RR, MCS and NCS algorithms. We consider the case of identical throughput guarantees for the MTGS algorithm. The expressions of ASSE for the RR, MCS and NCS algorithms are given in Appendix D. Figure 3.10 shows the ASSE for N Rayleigh-distributed users whose average CNRs are distributed with an average of 15 dB. The average CNR of the worst user is 5 dB and the best user has an average CNR of 17.79 dB. As expected, the MCS algorithm has the best ASSE performance. Since the NCS and the MTGS algorithms do not schedule the user with the absolute best channel for each time-slot, they have lower ASSE than MCS for these set of

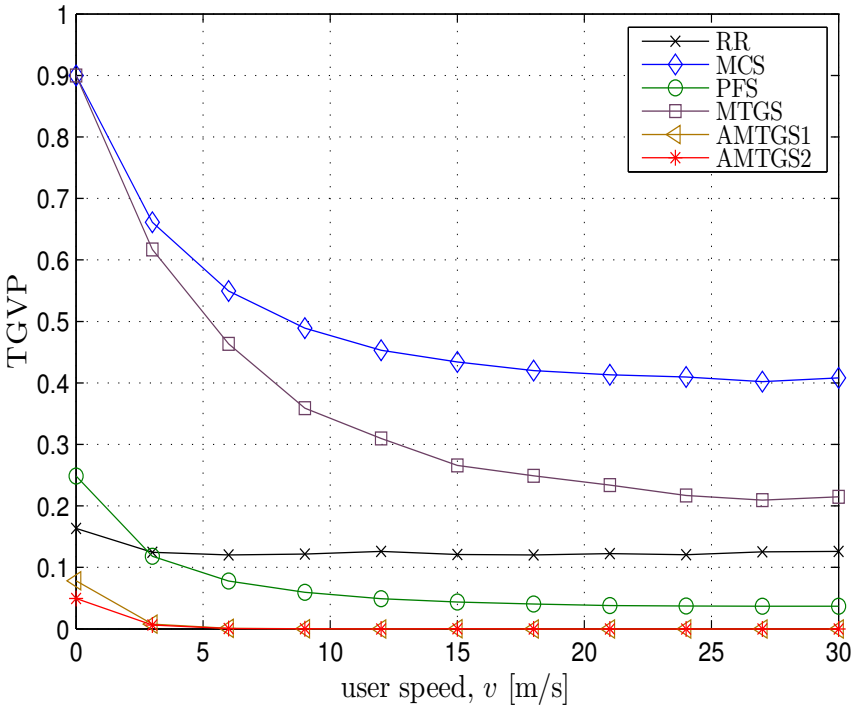


FIGURE 3.9: TGVP vs. Speed for 10 users in a Mobile WiMAX network with identical throughput guarantees of $B/(WT_W) = 0.2$ bits/s/Hz, where $T_W = 80$ ms corresponding to 16 time-slots. Each value in the plot is an average over 1000 Monte Carlo simulations.

users. The RR scheduling has the worst performance, since it schedules the users in a sequence irrespective of their channel conditions. The ASSE of the MTGS algorithm is worst for $N = 2$ because unlike other algorithms, it sets high access probability for the bad user, thus lowering the ASSE of the system.

3.9.5 Fairness Performance

In this section, we analyze the fairness of the MTGS algorithm, and the RR, MCS and NCS algorithms[8]. Figure 3.11 shows the time-slot fairness as a

[8]The expressions for time-slot and throughput fairness of the RR, MCS and NCS algorithms are given in Appendix E.

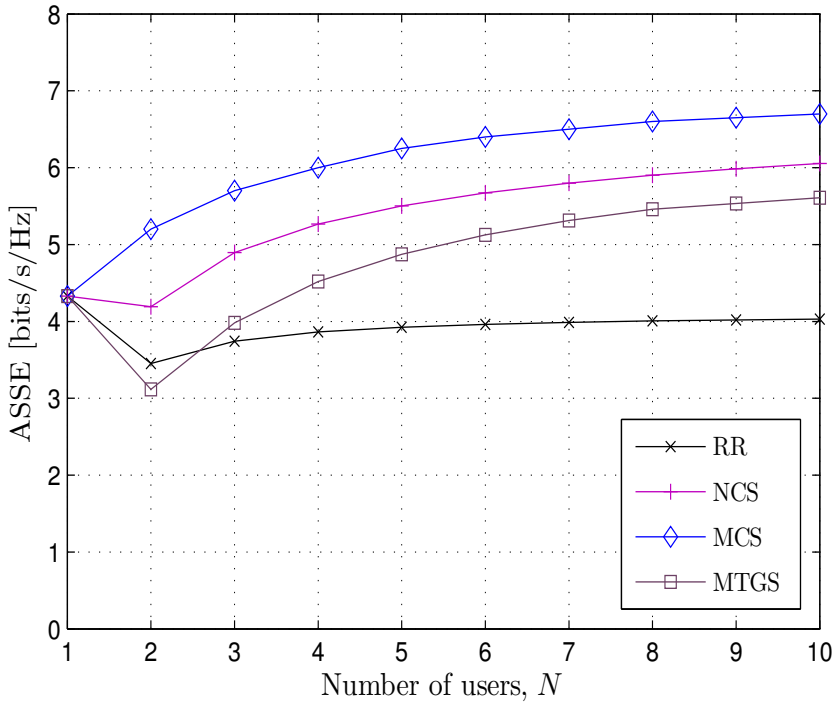


FIGURE 3.10: Comparison of ASSE for various scheduling algorithms, for N Rayleigh-distributed users whose average CNRs are distributed with an average of 15 dB.

function of the number of time-slots K for 10 users with Rayleigh fading channels given in Table 2.1. We see that the RR algorithm converges relatively fast to unity time-slot fairness. The NCS and MTGS algorithms also show very good time-slot fairness. However, the time-slot fairness of the MCS algorithm is the worst since it always allocates more time-slots to the user with the best channel.

The throughput fairness as a function of the number of time-slots K for different algorithms is shown in Figure 3.12. We again have 10 users with Rayleigh fading channels given in Table 2.1. It is interesting to note that the MTGS algorithm converges to the highest throughput fairness for large values of K. This is because this algorithm is designed to fulfill the throughput guarantees of all the users in the network. The throughput

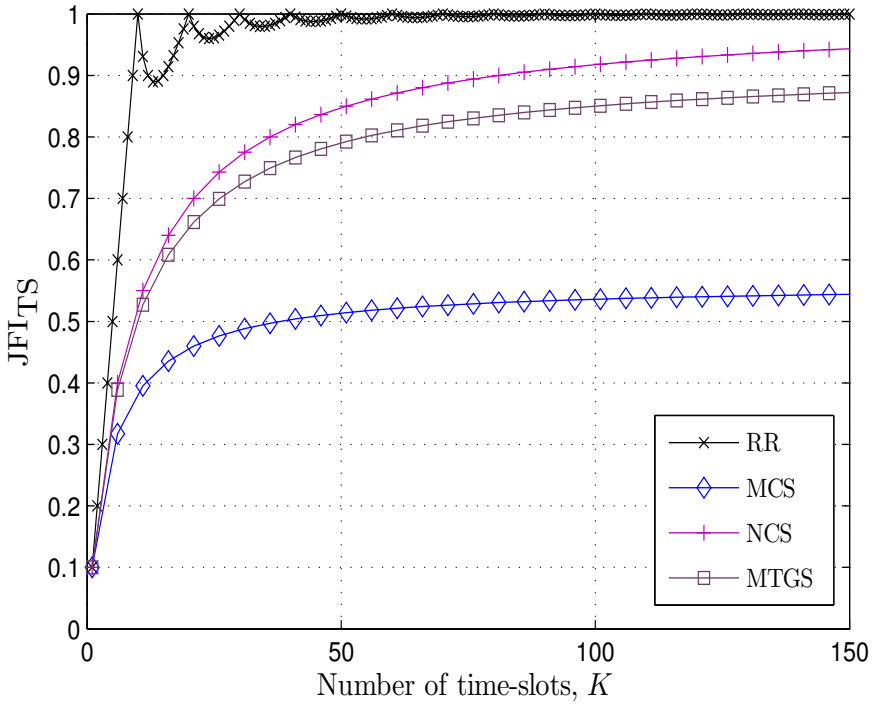


FIGURE 3.11: Time-slot fairness for 10 users with Rayleigh fading channels whose average CNRs are distributed with an average of 15 dB.

fairness of the MCS algorithm is again the worst since it allocates more time-slots to the best user and ignores the worst user.

3.10 Summary and Discussion

For wireless networks carrying real-time traffic, providing throughput guarantees is interesting both from the customers' and the network providers' point of view. In order to have the most efficient utilization of the network, a scheduler in such a network should try to distribute the amount of bits that can be received or transmitted by each user according to given throughput guarantees. In this chapter, we have formulated an optimization problem which aims at finding the maximum number of bits that can be guaranteed to the users within a time-window for a given set of system

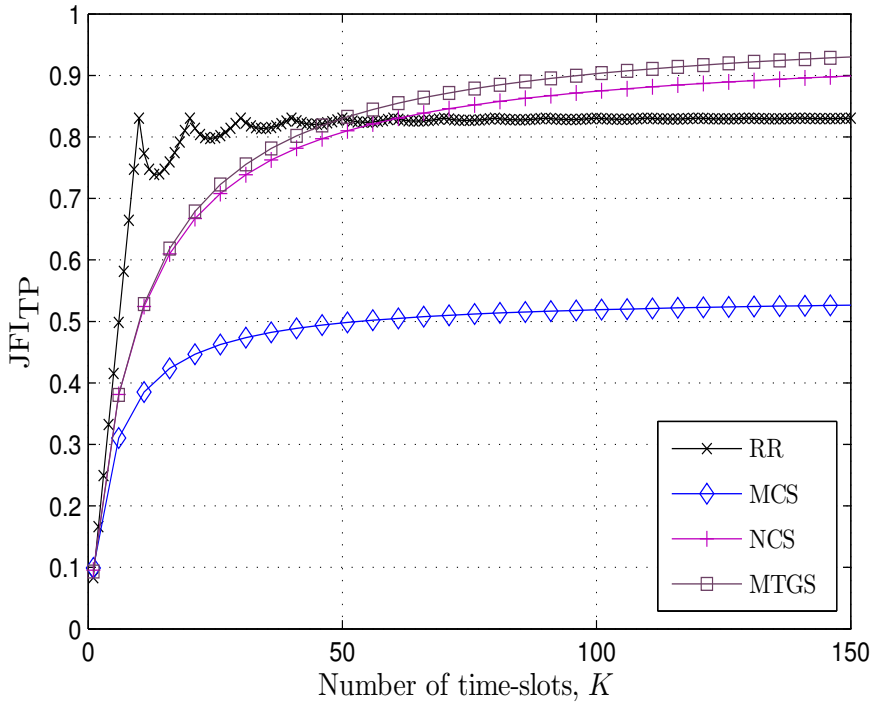


FIGURE 3.12: Throughput fairness for 10 users with Rayleigh fading channels whose average CNRs are distributed with an average of 15 dB.

parameters. By building on the results in [Borst and Whiting, 2003] and by assuming that the distributions of the users' CNRs are known, we have found an optimal scheduling algorithm (MTGS), both for the case where the throughput guarantees are different from user to user and for the case where the users have the same throughput guarantees. To further improve the short-term performance of this algorithm, we propose two adaptive and low complexity versions of the MTGS algorithm. The AMTGS2 algorithm is simpler in implementation but still provides similar throughput guarantees as provided by the AMTGS1 algorithm. Furthermore, it is also independent of the time-window and therefore overcomes the problem of fixed time-window placement of the AMTGS1 algorithm.

Results from our simulations show that the proposed adaptive algorithms perform significantly better than any of the other well-known schedul-

ing algorithms in networks based on Mobile WiMAX, HSDPA, LTE and WINNER I. For systems that have many time-slots within the time-window, e.g. for WINNER I, the MTGS algorithm also performs better than all the other well-known algorithms. For a network with heterogeneous throughput guarantees, the AMTGS1 and AMTGS2 algorithms are the only algorithms that support a throughput guarantee close to unity, i.e. TGVP = 0. The simulations further show that the performance of the adaptive algorithms at lower users' speed (strong temporal correlation) also remains the best. We have also analyzed the ASSE and fairness of the MTGS algorithm and showed that the throughput fairness of this algorithm is the best.

It should be noted that the analysis in this chapter involves several idealistic assumptions. For example, we assume that the CNR can be estimated perfectly and fed back with infinite precision and no delay, that ideal adaptive modulation and coding can be performed, that the CNR distributions of the users can be estimated perfectly, that the population of backlogged users is constant over the time-window the throughput guarantees are calculated, and that there is no capacity reduction due to protocol overheads. How realistic these assumptions are for real-life networks is a subject for further research and the next chapter.

Chapter 4

Effect of Imperfect Channel Knowledge on Throughput Guarantees

The analysis in Chapter 3 involves several idealistic assumptions, the most important of which is that each user estimates/predicts its CNR perfectly, and there is no feedback delay. In practice however, the channel prediction/estimation is not perfect, and there is always some delay in the feedback channel. If the data is encoded at a rate r [bits/s] that is higher than the maximum rate of reliable communication supported by the channel, then the system is said to be in *outage* [Tse and Viswanath, 2005]. The probability of outage for user i is given as

$$p_{out}(i) = \Pr[W \log_2(1 + \gamma_i) < r], \tag{4.1}$$

where γ_i is the actual CNR of user i.

In this chapter, we assume that a *maximum a posteriori* (MAP) predictor is employed [Holm, 2002; Øien, Holm, and Hole, 2004] for the channel coefficients, so that the system takes the feedback delay into account. We shall then investigate the effect of imperfect channel prediction and delay (prediction lag) on the throughput guarantees promised to all the users in the wireless network. Since imperfect channel prediction is assumed, the probability of outage cannot be zero. Therefore, we have also proposed a rate back-off mechanism to reduce the outage probability of the system to an acceptable level.

The rest of this chapter is organized as follows. We present the system model and describe the MAP-optimal predictor in Section 4.1. In Section 4.2, we propose a rate back-off mechanism to reduce the outage probability in case of imperfect channel knowledge. The optimization problem and the solution for obtaining the highest possible throughput guarantee in this case is discussed in Section 4.3. In Section 4.4, we describe the corresponding adaptive scheduling algorithm. We present the system parameters and numerical results in Sections 4.5 and 4.6 respectively, and give a brief summary in Section 4.7.

This chapter is based on [Rasool and Øien, 2012b].

4.1 System Model

We have a single base station that serves N backlogged users using TDM. Our analysis is valid both for the uplink and the downlink; in either case it is assumed that the total available bandwidth for the users is W [Hz], and that the users have constant transmit power.

We assume that the users experience Rayleigh faded channel gains. Each user i predicts his own CNR $\hat{\gamma}_i$ using a pilot-symbol assisted MAP-optimal predictor described in the next section. Before performing downlink scheduling the base station is assumed to receive these CNR measurements from all the users. The base station also performs uplink scheduling based on the predicted channel estimates, and for each time-slot, the base station takes a scheduling decision and distributes this decision to the selected user before transmission starts.

The communication channel between the base station and every one of the users is modeled by a flat, block-fading channel subject to AWGN noise. It is assumed that the communication channels corresponding to the different users fade independently. The block duration equals one time-slot, and is given by T_{TS} [seconds].

We also assume that the CNR values corresponding to different time-slots are correlated according to the modified Jakes' correlation model. Constant average CNR values for users are assumed in the time-window over which the throughput guarantees are calculated. Another important assumption is that the population of backlogged users is constant and equal to N. We have also assumed that only one user can be scheduled in a time-slot, and that the users always have data to send.

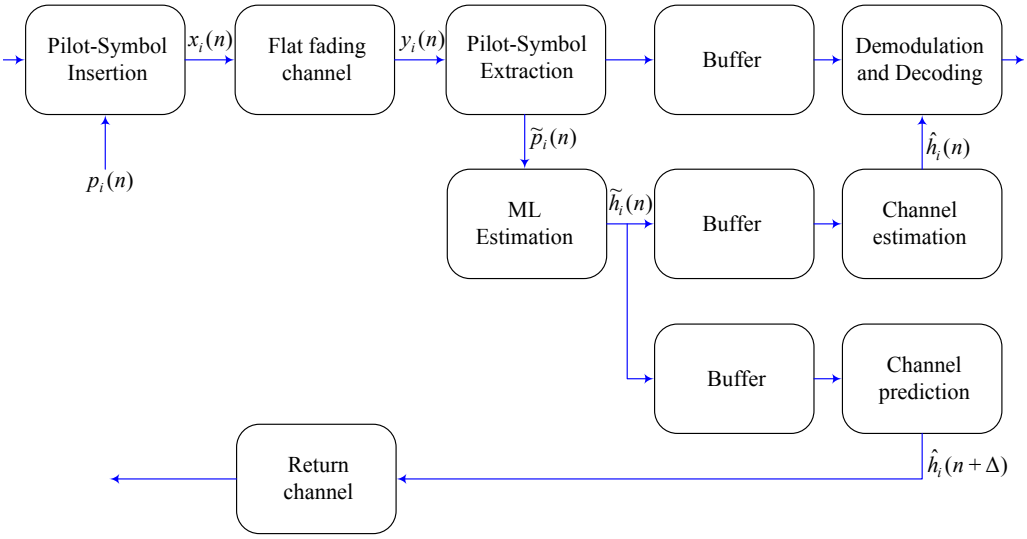


FIGURE 4.1: Pilot-symbol assisted modulation for channel prediction.

4.1.1 Pilot-Symbol Assisted MAP-Optimal Prediction

We consider the system described in Figure 4.1. Deterministic pilot-symbols, $p_i(n)$, are inserted into the data stream at time instants known to both the base station and user i. The symbols, $x_i(n)$, are transmitted on the communication channel described in Figure 1.2. The received signal, $y_i(n)$, is thus given as

$$y_i(n) = h_i(n) \cdot x_i(n) + n_i(n). \tag{4.2}$$

It is assumed that the pilot-symbols are inserted after equal spaced intervals, such that $x_i(n)$ is the data signal except at the pilot-symbol instants $n = l\mathcal{L}$, where $\mathcal{L}$ is the spacing between two adjacent pilot-symbols and l is an integer. At each pilot-symbol instant $n = l\mathcal{L}$, the following maximum likelihood (ML) estimate of $h_i(n)$ is calculated by the receiver [Øien et al., 2004]:

$$\begin{aligned} \tilde{h}_i(n) &= \frac{y_i(n)}{a_p} \\ &= h_i(n) + \frac{n_i(n)}{a_p}, \end{aligned} \tag{4.3}$$

where a_p is the amplitude of the pilot-symbols.

Channel prediction is done using $\mathcal{K}$ such ML estimates in the past to predict the CNR at the future symbol at the time instant $n+\Delta$ where $\Delta > 0$. The buffered estimates at time instant n can be written as

$$\tilde{\mathbf{h}}_i(n) = [\tilde{h}_i(n), \tilde{h}_i(n-\mathcal{L}), \ldots, \tilde{h}_i(n-(\mathcal{K}-1)\mathcal{L})]^{\mathbb{T}}, \tag{4.4}$$

where $\mathbb{T}$ is the transpose operator.

The predicted channel value can be written as

$$\hat{h}_i(n+\Delta) = \mathbf{c}_{\Delta}^{\mathbb{T}} \tilde{\mathbf{h}}_i(n), \tag{4.5}$$

where $\mathbf{c}_{\Delta} = [c_{\Delta}(0), c_{\Delta}(1), \ldots, c_{\Delta}(\mathcal{K}-1)]^{\mathbb{T}}$ is the vector containing the prediction filter coefficients which correspond to predicting the symbol at time instant $n+\Delta$. The MAP-optimal filter coefficient vector of length $\mathcal{K}$ on a Rayleigh fading channel is given as [Holm, 2002; Øien et al., 2004]

$$\mathbf{c}_{\Delta,\mathrm{MAP}}^{\mathbb{T}} = \mathbf{r}_{\Delta,\mathcal{K}}^{\mathbb{T}} \left(\mathbf{R}_{\mathcal{K}} + \frac{1}{\bar{\gamma}_i} \mathbf{I}_{\mathcal{K}\times\mathcal{K}} \right)^{-1}, \tag{4.6}$$

where $\mathbf{I}_{\mathcal{K}\times\mathcal{K}}$ is a $\mathcal{K} \times \mathcal{K}$ identity matrix. The vector $\mathbf{r}_{\Delta,\mathcal{K}}$ is of length $\mathcal{K}$ and its elements represent the correlation between the fading to be predicted at time $n+\Delta$ and the fading at the pilot-symbol instants. $\mathbf{R}_{\mathcal{K}}$ is the autocorrelation matrix of the fading at the pilot-symbol instants. With the assumption of Jakes spectrum in the fading process, the elements of $\mathbf{r}_{\Delta,\mathcal{K}}$ and $\mathbf{R}_{\mathcal{K}}$ can be calculated from the following equations [Øien et al., 2004]:

$$[\mathbf{r}_{\Delta,\mathcal{K}}]_l = R((\Delta + l\mathcal{L})T_s), \tag{4.7}$$

and

$$[\mathbf{R}_{\mathcal{K}}]_{lm} = R(|l-m|\,\mathcal{L}T_s), \tag{4.8}$$

where T_s is the symbol duration, l and m are integers, and $R(\tau)$ is the autocorrelation function given as

$$R(\tau) = J_0(2\pi f_D \tau), \tag{4.9}$$

where $J_0(x)$ is the zeroth-order Bessel function of the first kind.

The predicted CNR, $\hat{\gamma}_i(n+\Delta)$, at the time instant $n+\Delta$ is calculated from the predicted channel $\hat{h}_i(n+\Delta)$, and is given as

$$\hat{\gamma}_i(n+\Delta) = \frac{|\hat{h}_i(n+\Delta)|^2 P}{N_0 W}. \tag{4.10}$$

With a MAP-optimal predictor, the predicted CNR follows an exponential distribution with average CNR $\bar{\hat{\gamma}}_i = \rho_i \bar{\gamma}_i$ [Øien et al., 2004], where ρ_i is the normalized correlation between the actual and predicted CNRs, and $\bar{\gamma}_i$ is the actual average CNR of user i. The PDF, $f_{\hat{\gamma}_i}(\hat{\gamma})$, is then given as

$$f_{\hat{\gamma}_i}(\hat{\gamma}) = \frac{1}{\rho_i \bar{\gamma}_i} e^{-\frac{\gamma}{\rho_i \bar{\gamma}_i}}. \tag{4.11}$$

The normalized correlation, ρ_i, is given as [Øien et al., 2004]

$$\rho_i = \mathbf{r}_{\Delta,\mathcal{K}}^{\mathbb{T}} \left(\mathbf{R}_{\mathcal{K}} + \frac{1}{\bar{\gamma}_i} \mathbf{I}_{\mathcal{K}\times\mathcal{K}} \right)^{-1} \mathbf{r}_{\Delta,\mathcal{K}}. \tag{4.12}$$

We observe from (4.12) that ρ_i is a function of average CNR $\bar{\gamma}_i$, the Doppler spread f_D, the delay (prediction lag) ΔT_s, the filter length $\mathcal{K}$, and the pilot spacing $\mathcal{L}$.

4.2 Back-off CNR

For perfect channel prediction ($\rho_i = 1$), the predicted and the actual CNR are equal, i.e. $\hat{\gamma}_i = \gamma_i$, and there would be no outage if we send the data at the rate $\hat{r}_i$, where

$$\hat{r}_i = W \log_2(1 + \hat{\gamma}_i).$$

But due to imperfect channel prediction, there is always a chance that $\hat{\gamma}_i > \gamma_i$, and sending the data at the rate $\hat{r}_i$ would then result in outage. In this section, we therefore introduce a back-off CNR γ_b ($< \hat{\gamma}_i$), and send the data at a rate based on this back-off CNR. It should be noted that this would still not avoid the outage completely. It is, however, desirable to control the probability of this event, i.e. to demand

$$\Pr[\gamma_i < \gamma_b | \hat{\gamma}_i] \leq \epsilon, \tag{4.13}$$

where ϵ is a small constant chosen by the system designer. From [Jetlund, Øien, Holm, and Hole, 2004; Øien, Holm, and Hole, 2002], the probability of the complementary event is given as

$$\begin{aligned} 1 - \epsilon &= \Pr[\gamma_b \leq \gamma_i | \hat{\gamma}_i] \\ &= Q\left(\frac{\sqrt{\hat{\gamma}_i}}{\sqrt{\bar{\gamma}_i (1-\rho_i)/2}}, \frac{\sqrt{\gamma_b}}{\sqrt{\bar{\gamma}_i (1-\rho_i)/2}} \right), \end{aligned} \tag{4.14}$$

where

$$Q(a,b) = \int_b^\infty x I_0(ax) e^{-\frac{1}{2}(a^2+x^2)} dx \tag{4.15}$$

is the Marcum-Q function [Marcum, 1948]. The derivation of (4.14) is given in Appendix F. The back-off CNR γ_b is thus related to the predicted CNR $\hat{\gamma}_i$ through the Marcum-Q function given in (4.14).

To the authors' best knowledge the closed form expression for the inverse of the Marcum-Q function does not exist. But as suggested in [Jetlund et al., 2004], the inverse of one of the arguments can be obtained, with a given accuracy, using e.g. Ridders' method [Ridders, 1979][1]. That is, the inverse of $Q(\cdot,\cdot)$ with respect to $\hat{\gamma}_i$ can then be written as

$$\gamma_b(\epsilon) = \begin{cases} \left(\sqrt{\frac{\hat{\gamma}_i(1-\rho_i)}{2}} \cdot q_b\left(\frac{\sqrt{\hat{\gamma}_i}}{\sqrt{\frac{\hat{\gamma}_i(1-\rho_i)}{2}}}, \epsilon \right) \right)^2, & \rho_i < 1 \\ \hat{\gamma}_i, & \rho_i = 1 \end{cases} \tag{4.16}$$

where $b = q_b(a,\epsilon)$ is the inverse of the complementary Marcum-Q function, $1 - Q(a,b)$, with respect to its second argument.

Using the back-off CNR, the instantaneous rate of user i will be given as

$$r_i = \begin{cases} W \log_2(1+\gamma_b(\epsilon)) & \text{if } \gamma_b(\epsilon) \leq \gamma_i \\ 0 & \text{if } \gamma_b(\epsilon) > \gamma_i \end{cases}. \tag{4.17}$$

The average rate $\bar{R}_i$ for user i when he is transmitting or receiving, can then be written as

$$\bar{R}_i = W \int_0^\infty \log_2(1+\gamma_b(\epsilon)) f_{\hat{\gamma}*}(\hat{\gamma}|i) \Pr[\gamma_b \leq \gamma_i | \hat{\gamma}_i] d\hat{\gamma}, \tag{4.18}$$

where $f_{\hat{\gamma}*}(\hat{\gamma}|i)$ is the PDF of the predicted CNR of user i when this user is scheduled. Just like in Section 3.2, we have assumed that the time-window T_W is long enough and contains enough time-slots for the channel to reveal its ergodic properties, and that the Shannon capacity can be achieved.

4.3 The Optimization Problem and its Solution

The optimization problem in this chapter, under the assumption of identical throughput guarantees for all the users, builds on the optimization

[1] MATLAB implementation of the Ridders' method is available in [Kiusalaas, 2010].

problem described in Section 3.2. The objective is thus to find a *scheduling algorithm* that maximizes the throughput guarantee B promised to all the users over the time-window T_W, under the assumption of imperfect channel prediction. This means that (3.3) has to be maximized subject to the constraints in (4.18), for $i = 1, \ldots, N$.

Following Section 3.3, the scheduling algorithm that solves the problem described above is given as

$$i^*(n) = \underset{1 \le i \le N}{\operatorname{argmax}} \left(\frac{r_i(n)}{\alpha_i} \right), \tag{4.19}$$

where $i^*(n)$ is the index of the user that is going to be scheduled in time-slot n, $r_i(n)$ is the instantaneous rate of user i in time-slot n given by (4.17), and α_i is a constant to be optimized.

To find the optimal parameters α_is for this algorithm, however, we need to know the PDF of $\gamma_b(\epsilon)$. This is not possible since we do not have a closed form expression for $\gamma_b(\epsilon)$, as mentioned in the previous section. However, we know the PDF of the predicted CNR $\hat{\gamma}_i$ (given in (4.11)), and suggest a two-step alternate (sub-optimal) solution based on that PDF. These steps are *user selection* and *rate calculation*, which are discussed below.

4.3.1 User Selection

At the time-slot n, the user who is going to be scheduled is now given as

$$i^*(n) = \underset{1 \le i \le N}{\operatorname{argmax}} \left(\frac{\hat{r}_i(n)}{\alpha_i} \right), \tag{4.20}$$

i.e.,

$$r_i(n) = \hat{r}_i(n) = W \log_2(1 + \hat{\gamma}_i(n)),$$

where $\hat{\gamma}_i(n)$ is the predicted CNR of user i at time-slot n.

To find the α_is, we proceed as in Section 3.3 and again define the random variable $S_i \triangleq \frac{R_i}{\alpha_i}$, where R_i is the random variable describing the rate of user i. For flat, block-fading AWGN channels, the maximal value of the metric S_i for user i within a time-slot (block) with CNR $\hat{\gamma}$ can thus be expressed as

$$S_i(\hat{\gamma}) = \frac{W \log_2(1 + \hat{\gamma})}{\alpha_i}. \tag{4.21}$$

The PDF and the CDF for the normalized rate $S_i = s$ for user i are then given as

$$\begin{aligned} f_{S_i}(s) &= \left.\frac{f_{\hat{\gamma}_i}(\hat{\gamma})}{\frac{dS_i(\hat{\gamma})}{d\hat{\gamma}}}\right|_{\hat{\gamma}=2^{\frac{s\cdot\alpha_i}{W}}-1} \\ &= \frac{\alpha_i \ln(2)}{W\rho_i\bar{\gamma}_i} 2^{\frac{s\cdot\alpha_i}{W}} e^{-\frac{2^{\frac{s\cdot\alpha_i}{W}}-1}{\rho_i\bar{\gamma}_i}}, \end{aligned} \tag{4.22}$$

and

$$\begin{aligned} F_{S_i}(s) &= \int_0^s f_{S_i}(x)dx \\ &= 1 - e^{-\frac{2^{\frac{s\cdot\alpha_i}{W}}-1}{\rho_i\bar{\gamma}_i}} \end{aligned} \tag{4.23}$$

respectively. Inserting (4.22) and (4.23) in (3.12)-(3.14), and then combining (3.5), (3.12), and (3.14) we again obtain $3N$ equations in $3N$ unknowns, and can thus find the values for the α_is.

4.3.2 Rate Calculation

After selecting the "*best*"[2] user according to (4.20), data will then be transmitted to this user at a rate calculated on the basis of the back-off CNR $\gamma_b(\epsilon)$, given in (4.16). As mentioned before, this will still not avoid the outage completely. However, this back-off will ensure that the outage probability of the system is reduced to an acceptable level given by ϵ.

4.4 Improving the Short-Term Performance

Similar to the MTGS algorithm in Chapter 3, the scheduling algorithm in the previous section is designed for long time-windows containing many time-slots. For shorter time-windows, two adaptive scheduling algorithms were proposed in Section 3.5. In this chapter, we shall make use of the AMTGS2 algorithm given in Section 3.5.2 to analyze the system performance. This algorithm for the case of imperfect channel prediction is given as:

[2]It should be noted that this is the best user based on imperfect channel prediction. But if the channel prediction is incorrect, the selected user may not be the best in reality.

For promised throughput guarantees B, select a user $i^*(n)$ that has a maximum

$$i^*(n) = \underset{1\leq i\leq N}{\operatorname{argmax}} \left(v_i(n-1)\frac{\hat{r}_i(n)}{\alpha_i} \right), \tag{4.24}$$

where $v_i(n)$ is given in (3.22).

4.5 System/Simulation Parameters

For the simulations in the next section, we assume

- carrier frequency $f_c = 1$GHz,
- user speed $v = 30$ m/s,
- filter length $\mathcal{K} = 250$, and
- pilot spacing $\mathcal{L} = 10$.

The results are shown for 10 users having Rayleigh fading channels, given in Table 2.1. Without loss of generality, the user indices are assigned such that user 1 has the lowest average CNR and user 10 has the highest average CNR.

For the Mobile WiMAX system, the time-slot length T_{TS} and symbol duration T_s for the downlink are 5 ms and 102.9 μs [WiMAX Forum, 2006], respectively. The European IST research project WINNER I has suggested a time-slot duration of 0.34 ms and a symbol duration of 21.8 μs for a future wireless system [Sternad et al., 2006]. The corresponding time-slot length and the symbol duration for the 3GPP LTE network are 1 ms and 71.4 μs [Freescale Semiconductor, 2007], respectively. If we assume that $T_W = 80$ ms, the time-window contains 235, 80, and 16 time-slots for WINNER I, LTE, and Mobile WiMAX, respectively.

Each value in the plots is an average over 1000, 500, and 250 Monte Carlo simulations for the Mobile WiMAX, LTE and WINNER I systems respectively.

4.6 Numerical Results

As mentioned before in this chapter, the outage probability in the case of imperfect CSI cannot be completely reduced to zero. We therefore introduced a back-off CNR which was calculated based on a preset value of the

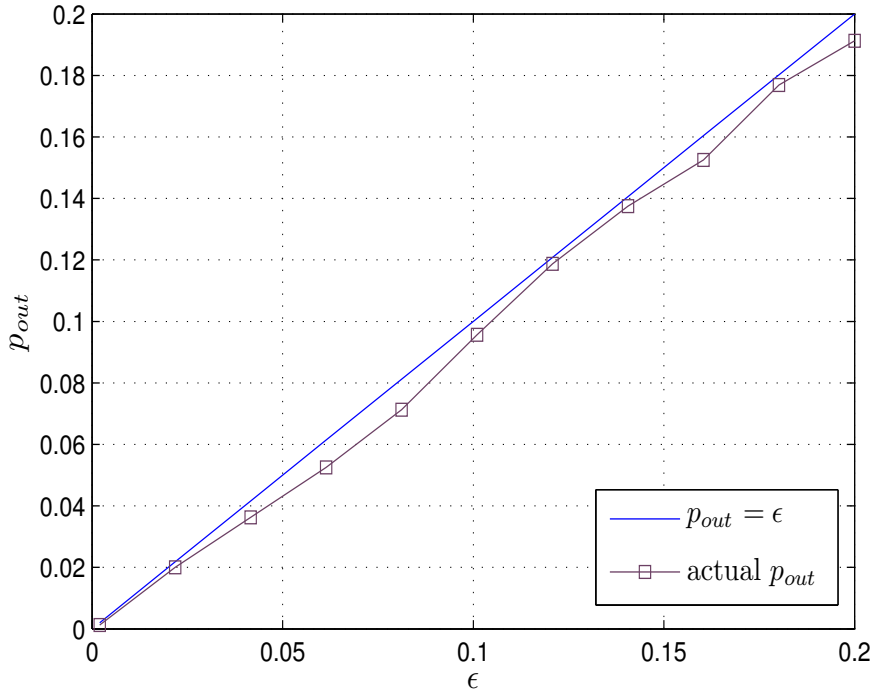


FIGURE 4.2: Comparison of the actual outage probability, based on simulations, with the chosen ϵ. The MTGS algorithm is considered, for 10 users in a Mobile WiMAX network. Each value in the plot is an average over 1000 Monte Carlo simulations.

outage probability, i.e. ϵ. In Figure 4.2, we plot the actual outage probability experienced in a 10 user Mobile WiMAX network for different values of ϵ, and for identical throughput guarantees $B/(WT_W) = 0.2$ bits/s/Hz, and $\Delta = 10$. The figure shows that the experienced outage probability, based on simulations, is actually equal to or below the chosen ϵ. Hence, the back-off mechanism indeed reduces the outage probability to an acceptable level given by ϵ.

In Figure 4.3, percentage loss in the average transmitted and received rates due to imperfect channel prediction, for different values of ϵ (chosen outage probability) is plotted. We again consider the 10 user Mobile WiMAX network with identical throughput guarantee $B/(WT_W) = 0.2$ bits/s/Hz, and $\Delta = 10$. The lower the value of ϵ, the lower will be the

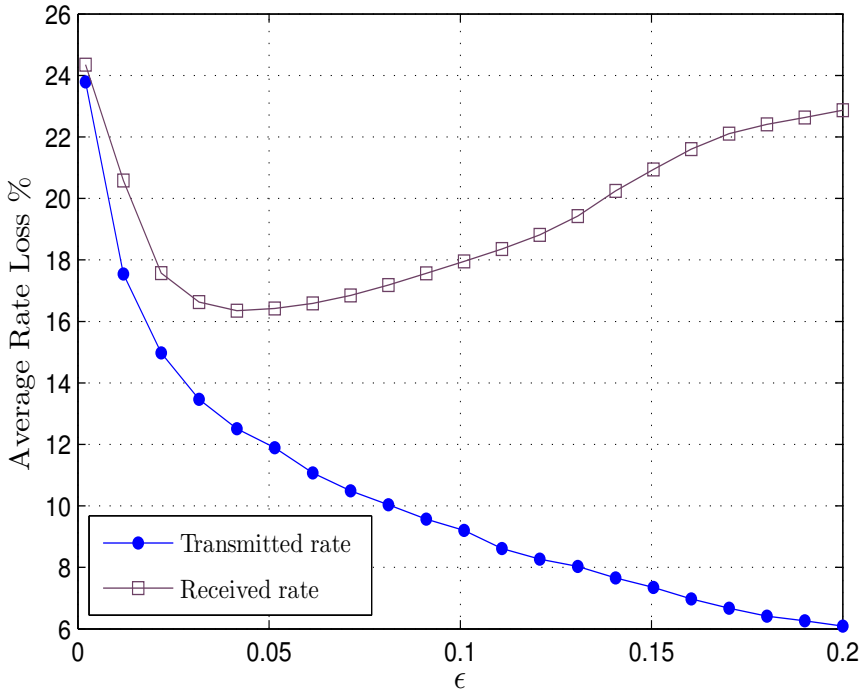


FIGURE 4.3: Average rate loss for different values of ϵ, plotted for 10 users in a Mobile WiMAX network using the MTGS algorithm. Each value in the plot is an average over 1000 Monte Carlo simulations.

back-off CNRs and the data rates, and the higher will be the rate loss. As we increase the value of ϵ, we observe a reduction in the rate loss. For higher values of ϵ, the data will be transmitted at higher rates. Therefore, the loss in the transmitted rate will be lower. But since the outage probability is now high, much of the data will be lost, reducing the average received rate, and thus the loss in the average received rate is higher. This is also evident from Figure 4.4 where we have plotted the TGVP for the MTGS algorithm, for different values of ϵ. From the figure, we observe that the TGVP is higher for higher values of ϵ, since the received rate loss is higher in this case. The TGVP for very low values of ϵ is also high. Since in that case, we send the data at very low rates in order to avoid outage, and due to low rates the system was not able to fulfill the throughput guaran-

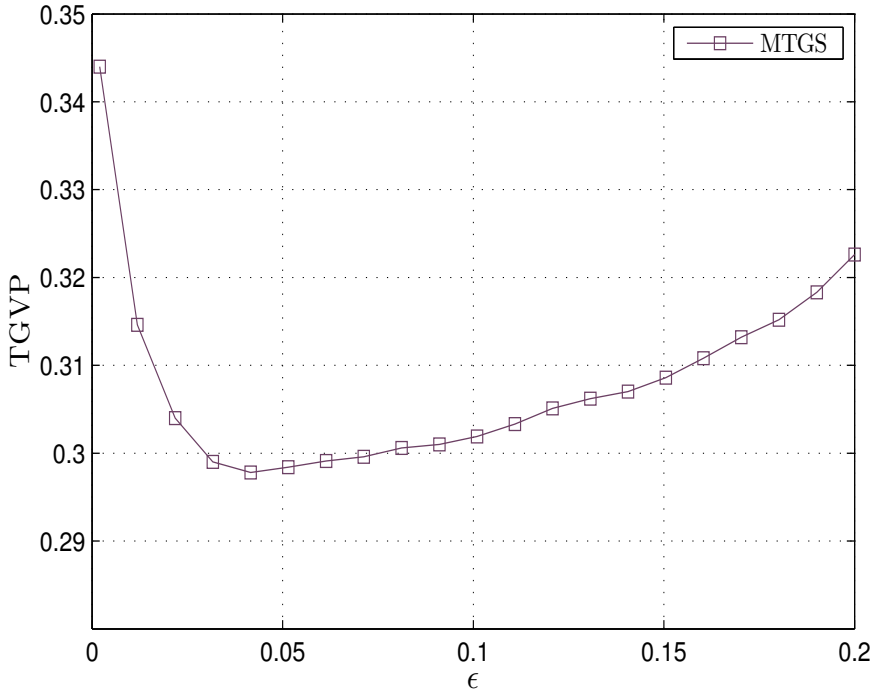


FIGURE 4.4: TGVP for MTGS algorithm for different values of ϵ, for $B/(WT_W) = 0.2$ bits/s/Hz, in a 10 user Mobile WiMAX network. Each value in the plot is an average over 1000 Monte Carlo simulations.

tees of all the users in the network. The TGVP is lowest for $\epsilon = 0.0416$ for this system. For the remaining plots, we will only consider this value of ϵ.

In Section 3.9, it is shown that for perfect channel prediction, the AMTGS2 algorithm described in Section 4.4 outperforms other scheduling algorithms in terms of the TGVP. So for the remaining plots in this section, we focus only on the performance of this scheduling algorithm under imperfect channel prediction. Furthermore, we only consider the scenario where all the users are promised identical throughput guarantees B/WT_W.

Figures 4.5-4.7 show the TGVP performance in networks that are respectively based on Mobile WiMAX, LTE and WINNER I. For the system parameters given in the previous section, we vary the time delay ΔT_s and observe the TGVP performance. From the figures, we observe that as Δ in-

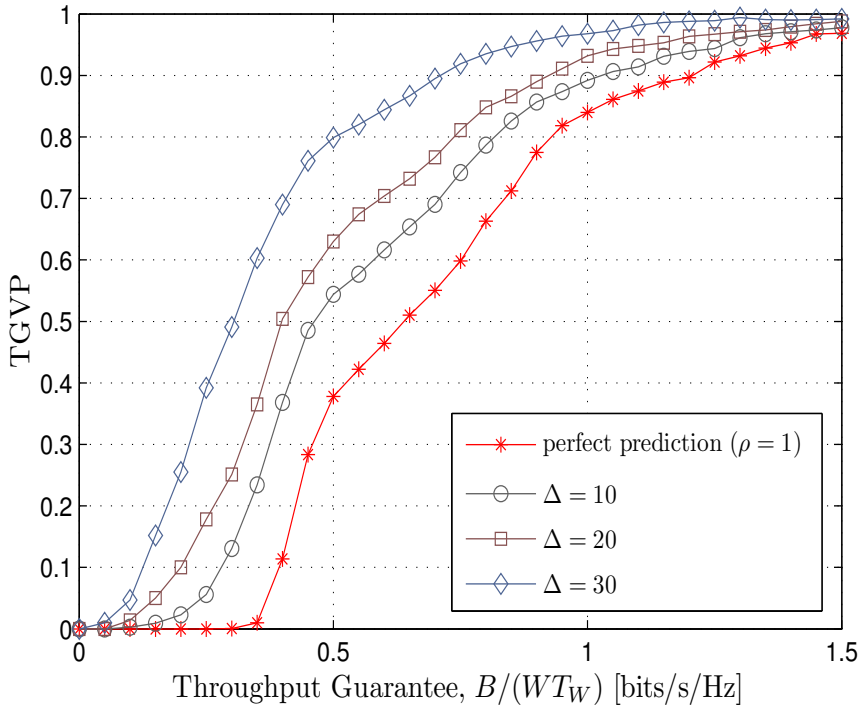


FIGURE 4.5: Effect of imperfect channel knowledge on the TGVP performance of AMTGS2 algorithm in a Mobile WiMAX network. Plotted for a time-window $T_W = 80$ ms that contains 16 time-slots. Each value in the plot is an average over 1000 Monte Carlo simulations.

creases, the TGVP performance of all the systems deteriorates. The higher the time delay, the lower will be the normalized correlation between the actual and predicted CNRs. For example, the value of the normalized correlation for the user 10 is $\rho_{10} = 0.98$ for $\Delta = 10$ and $\rho_{10} = 0.91$ for $\Delta = 30$ in the LTE system. We observe (as expected) that lower normalized correlation results in the deterioration of the system performance in terms of the maximum throughput that can be guaranteed. The performance loss in the case of Mobile WiMAX is the highest, since its symbol duration T_s is also the highest, causing a higher time delay or prediction lag ΔT_s, and hence the lowest normalized correlations for the users as compared to the LTE or WINNER I based systems. This can be seen in Figure 4.8, where the aver-

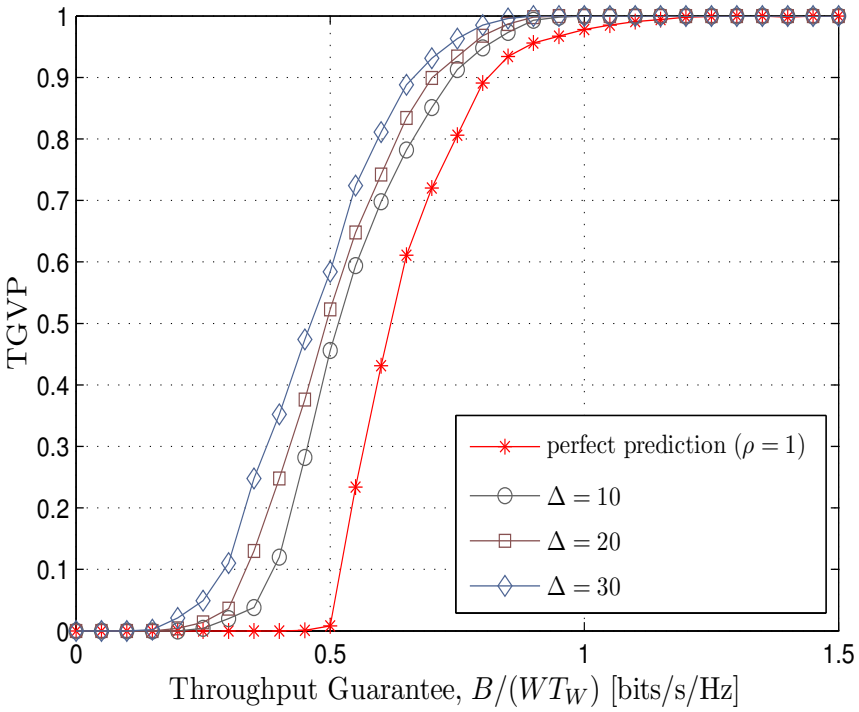


FIGURE 4.6: Effect of imperfect channel knowledge on the TGVP performance of AMTGS2 algorithm in an LTE network. Plotted for a time-window $T_W = 80$ ms that contains 80 time-slots. Each value in the plot is an average over 500 Monte Carlo simulations.

age normalized correlation ρ_{avg} (average over all the users) is plotted as a function of Δ, for the system parameters given in Section 4.5. We observe from the figure that the average normalized correlation for the WINNER I system does not change too much, whereas the average normalized correlation for the Mobile WiMAX system decreases significantly with increasing value of Δ. We can thus confirm what intuition tells us that for a given set of system parameters, it is better not to predict too far ahead in time when the symbol duration is large.

For the system parameters given in Section 4.5, the received average rate loss for Mobile WiMAX system turned out to be 16.351% for $\epsilon = 0.0416$, and $\Delta = 10$ (see Figure 4.3). The corresponding values of the

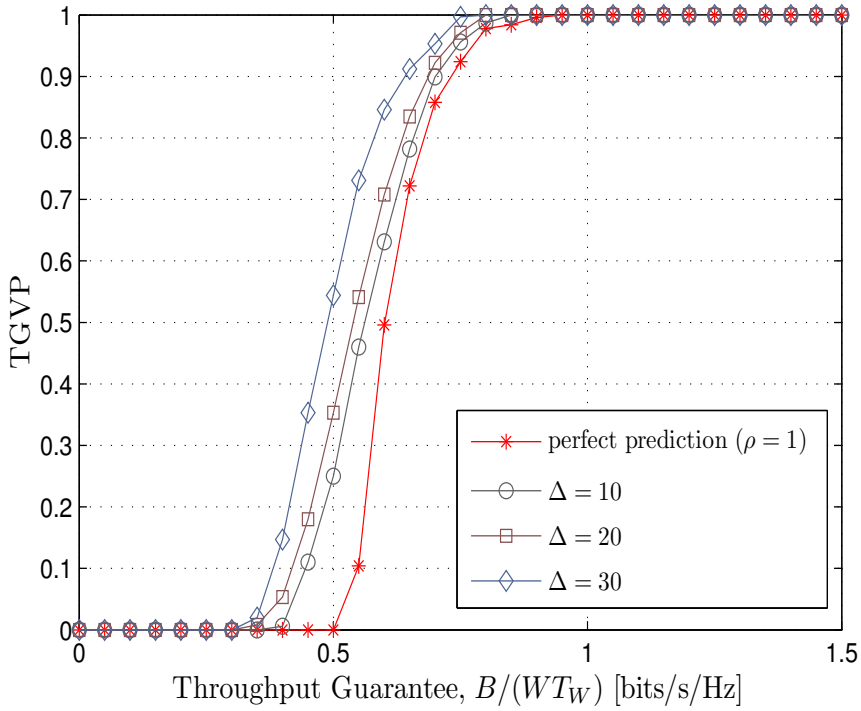


FIGURE 4.7: Effect of imperfect channel knowledge on the TGVP performance of AMTGS2 algorithm in WINNER I network. Plotted for a time-window $T_W = 80$ ms that contains 235 time-slots. Each value in the plot is an average over 250 Monte Carlo simulations.

normalized correlation for the worst and the best users are $\rho_1 = 0.77$ and $\rho_{10} = 0.96$ respectively. If we want to reduce the average rate loss to say 13.5%, we have found after various experiments that we need to improve the normalized correlation of the worst user to atleast $\rho_1 = 0.85$. There will be an improvement in the normalized correlation of the remaining users also since it is assumed that they also use the same set of system parameters. This improvement can be achieved by changing either the filter length $\mathcal{K}$, the pilot spacing $\mathcal{L}$, or the Doppler frequency shift f_D. Thus we have an infinite number of possibilities to set the desired rate loss in our system. An example of the "possibility space" given that a maximum of 13.5% rate loss is acceptable is shown in Figure 4.9. We need to be on or below this

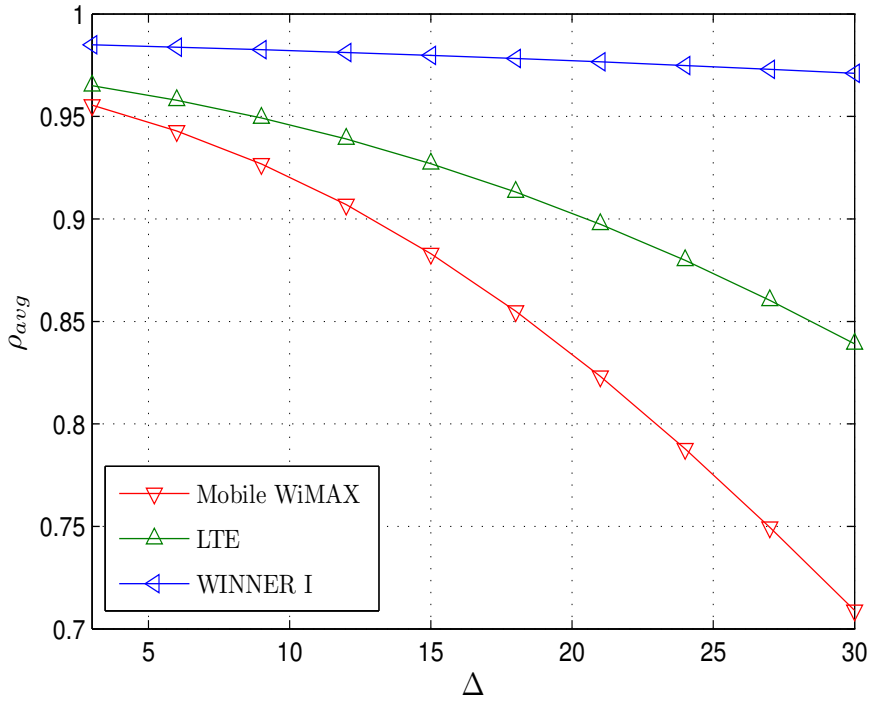


FIGURE 4.8: The average normalized correlation plotted as a function of Δ. Rest of the system parameters are given in Section 4.5.

surface in order to have maximum 13.5% received average rate loss for the Mobile WiMAX system. Ideally, we would like to have a larger value of $\mathcal{L}$ to minimze the rate loss due to pilot-symbols. We observe from the figure that for our system, increasing the filter length is not that beneficial in this regard. Therefore, if we want to keep $\mathcal{L} = 10$, we need to reduce the Doppler frequency shift to about 60 Hz to have a rate loss of 13.5%.

4.7 Summary and Discussion

In Chapter 3, we described an optimization problem and its solution which aims at maximizing the throughput guarantees offered to the users in a wireless network. However, that analysis assumes that each user estimates its CNR perfectly, and there is no feedback delay. This is not true in prac-

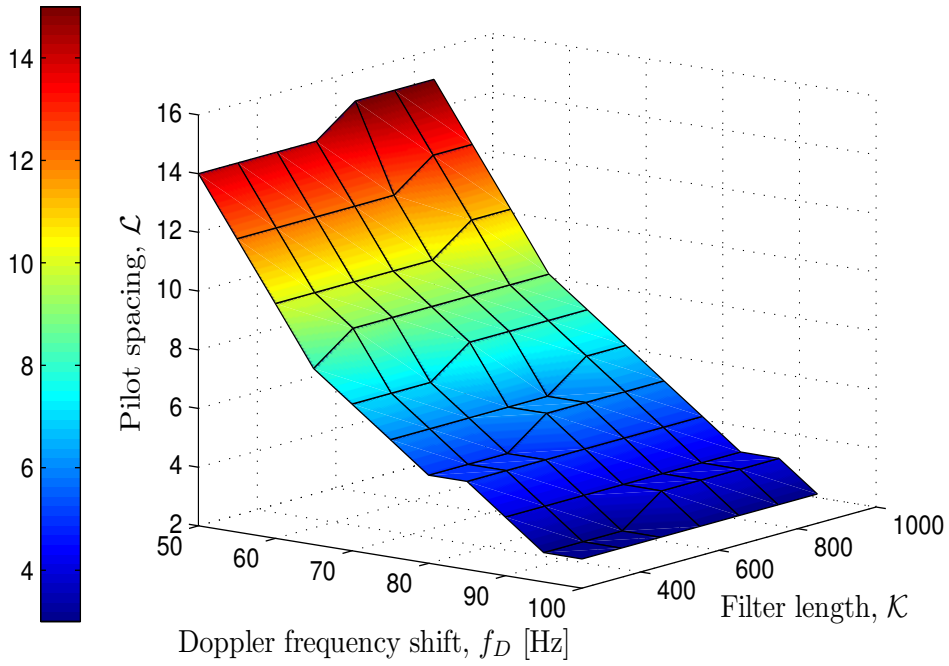


FIGURE 4.9: An example of the "possibility space" given that a maximum of 13.5% average rate loss, due to imperfect channel prediction, is acceptable in a Mobile WiMAX network. Plotted for a time-window $T_W = 80$ ms that contains 16 time-slots.

tice. Therefore in this chapter we have made use of a maximum a posteriori (MAP) predictor to take the delay (prediction lag) and channel noise into account. We have then investigated the effect of imperfect channel prediction and delay on the throughput guarantees promised to all the users in the wireless network. We have also proposed a rate back-off mechanism to reduce the outage probability of the system. It should be noted that the probability of outage cannot be completely reduced to zero. From the simulations, we observe that the TGVP performance of the system deteriorates when the normalized correlation between the actual and predicted CNR decreases. We also conclude from the performance comparison of Mobile WiMAX with LTE and WINNER I systems that it is better not to predict too far ahead in time when the symbol duration is large. Finally, we also

explored the "possibility space" given that a maximum of X% rate loss is acceptable.

Chapter 5

Improved Throughput Guarantees in MIMO Broadcast Channels

Research has shown that MIMO systems have great potential to achieve high throughput in wireless systems. Point-to-point multiple-antenna systems offer, under certain assumptions, an increase in the capacity which is proportional to the minimum of the number of transmit and receive antennas [Foschini and Gans, 1998], and we also see similar capacity gains in the case of multi-user MIMO communications [Tse and Viswanath, 2005, p. 254]. In multi-user systems, the achievable capacity gains are limited by the number of receive antennas at each user. Since the multiple-antenna gains with TDMA are very limited [Jindal and Goldsmith, 2005], the base station should send data to multiple users simultaneously in order to achieve the capacity gains. One multi-user strategy is opportunistic beamforming [Viswanath et al., 2002], the main principle of which is to randomly vary the complex weights at an antenna array, and to schedule the users with highest CNR for data transmission. In [Sharif and Hassibi, 2005], the authors extended this idea to multiple random beams. Their scheme generates M orthonormal random beams and schedules M out of N users with best signal-to-interference-plus-noise ratios (SINR) for those beams.

In this chapter, we extend the work in Chapter 3 to MIMO broadcast channels so that multiple users can be selected in a single time-slot, and propose scheduling algorithms that make use of orthonormal random beam-

forming, and try to fulfill the throughput guarantees promised to all the mobile users in a MIMO broadcast channel. We consider two beamforming scenarios in this chapter. In the first scenario, random beams are used for all the selected users, whereas transmit beamforming is used for the first selected user in the second scenario. We also describe corresponding adaptive scheduling algorithms that perform significantly better than other well-known scheduling algorithms, in terms of the TGVP.

The rest of this chapter is organized as follows: The system model is presented in Section 5.1. In Section 5.2, we formulate the optimization problem for obtaining the highest possible throughput guarantees over a time-window for the MIMO broadcast channel. In Section 5.3, we show how the solution to this problem can be found for the two beamforming scenarios mentioned above. We describe the corresponding adaptive scheduling algorithms in Section 5.4. Simulation parameters are given in Section 5.5. We present our numerical results in Section 5.6, and give a brief summary in Section 5.7.

This Chapter is based on [Rasool and Øien, 2011b,c].

5.1 System Model

We consider a single base station, equipped with M antennas, that serves N backlogged users, each having a single antenna. The total available bandwidth for the users is W [Hz], and the total transmit power is P. We further assume equal power allocation over each transmit beam, i.e. P/M. The corresponding MIMO broadcast channel model is shown in Figure 5.1.

The received signal of user i in this case, denoted by y_i, is given as[1]

$$y_i = \mathbf{h}_i^H \mathbf{W}\mathbf{x} + n_i, \tag{5.1}$$

where $\mathbf{h}_i$ is an $M \times 1$ complex Gaussian channel vector (with zero mean and variance σ_i^2) between the base station and user i, $\mathbf{W} = [\mathbf{w}_1 \mathbf{w}_2 \dots \mathbf{w}_M]$ is a beamforming matrix where $\mathbf{w}_m$ is an $M \times 1$ beamforming vector, n_i is an i.i.d. complex AWGN noise with zero mean and variance $\sigma^2 = N_0 W$, and $\mathbf{x} = [x_1, x_2, \dots, x_M]^T$ is an $M \times 1$ transmitted signal such that

$$\mathbb{E}[\mathbf{x}^H \mathbf{x}] = P.$$

[1] We have not used the time indices for simplification.

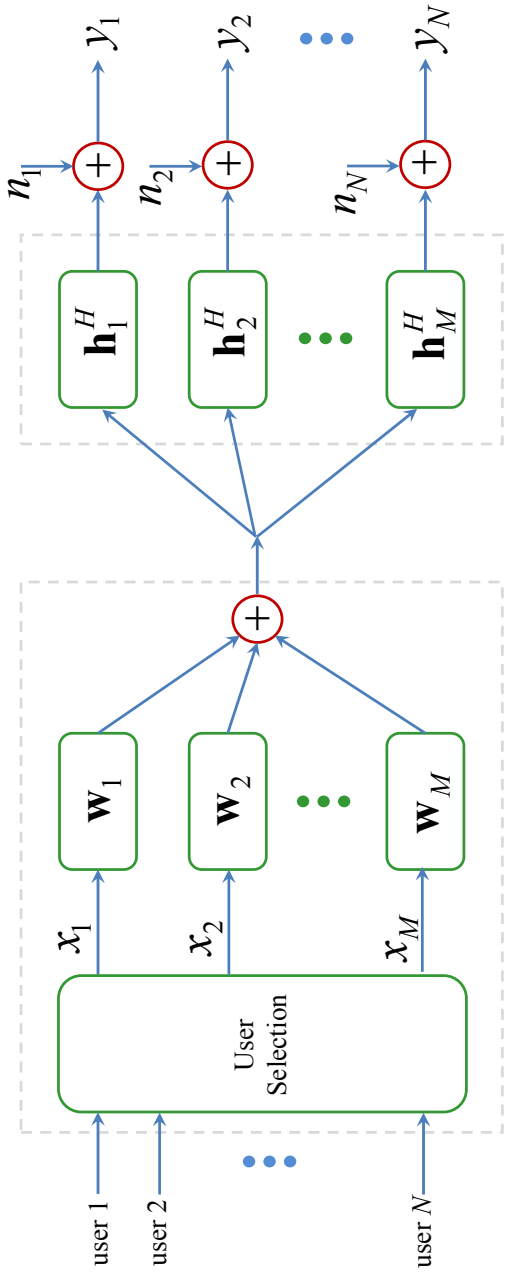


FIGURE 5.1: MIMO broadcast channel modeling our scenario

(5.1) can also be written as

$$\begin{aligned} y_i &= \sum_{j=1}^{M} \mathbf{h}_i^H \mathbf{w}_j x_j + n_i \\ &= \mathbf{h}_i^H \mathbf{w}_i x_i + \sum_{\substack{j=1 \\ j \neq i}}^{M} \mathbf{h}_i^H \mathbf{w}_j x_j + n_i. \end{aligned} \tag{5.2}$$

The SINR of user i, denoted by η_i, is given as

$$\eta_i = \frac{\left|\mathbf{h}_i^H \mathbf{w}_i\right|^2}{\sum_{j \neq i} \left|\mathbf{h}_i^H \mathbf{w}_j\right|^2 + M/\bar{\gamma}_i}, \tag{5.3}$$

where $\bar{\gamma}_i$ is the average received CNR of user i.

We also assume constant average CNR values for the time-window over which the throughput guarantees are calculated. Another important assumption is that the population of backlogged users during the time-window is constant and equal to N. We further assume that $N \geq M$, so that all the beams are utilized by different users at each time instance.

5.2 The Optimization Problem

We now formulate an optimization problem aimed at obtaining the maximal throughput guarantee B [bits], which can be achieved within a time-window of T_W [seconds] in a MIMO broadcast channel with M base station antennas and N single antenna users. This optimization problem builds on the optimization problem given in Section 3.2.

We again assume identical throughput guarantees promised to all the users in our system. In this case, the accumulated time of user i over the time-window T_W and the average rate for user i when he is scheduled, denoted by T_i [seconds] and $\bar{R}_i$ [bits/s] respectively, should again satisfy (3.1).

Since the number of users scheduled in a single time-slot is M, the sum of the T_is satisfies

$$\sum_{i=1}^{N} T_i = MT_W. \tag{5.4}$$

In our system, user i is allocated no more than one beam per time-slot. Therefore, $T_i \leq T_W$ for all $i = 1, ..., N$. Inserting (3.1) into (5.4), we get

$$B = \frac{MT_W}{\sum_{i=1}^{N} \frac{1}{\bar{R}_i}}. \tag{5.5}$$

Under the assumption that T_i is long enough to make the time-window T_W appear infinitely long, (3.1) can also be written as

$$p(i)MT_W\bar{R}_i = B, \tag{5.6}$$

where $p(i)$ is the access probability for user i within the duration of the time-window T_W. Setting (5.5) equal to (5.6), we again obtain (3.5).

The average rate $\bar{R}_i$ for user i when he is scheduled, can be written as

$$\bar{R}_i = W \int_0^\infty \log_2(1+\eta) f_{\eta*}(\eta|i) d\eta, \tag{5.7}$$

where $f_{\eta*}(\eta|i)$ is the PDF of the SINR of user i when this user is scheduled. As mentioned before in previous chapters, this requires that T_W is long enough for the channel to reveal its ergodic properties, and that the Shannon capacity is achieved.

Now, our goal is to find a *scheduling scheme* that gives the maximum B that can be promised to all the users over the time-window T_W, meaning that (5.5) has to be maximized subject to the constraints (5.7), for all $i = 1, \ldots, N$.

5.3 Solution to the Optimization Problem

The solution to the optimization problem discussed in the previous section depends on the beamforming scheme used in the system. As mentioned before, we consider two beamforming schemes in this section. In the first scheme, random beams are used for all the selected users, whereas transmit beamforming is used for the first selected user in the second scheme.

5.3.1 Random Beams for All the Selected Users

The beamforming strategy that we use in this subsection is similar to the one described in [Sharif and Hassibi, 2005]. The base station generates M orthonormal random beam vectors $\mathbf{w}_m$ according to isotropic distribution,

so that the beamforming matrix is $\mathbf{W} = [\mathbf{w}_1\mathbf{w}_2 \dots \mathbf{w}_M]$. This matrix can either be generated for every time-slot or can be generated once for a given time-window. By assuming that x_m is the desired signal and the other x_js are interference, the following M SINRs are computed for the user i:

$$\eta_{i,m} = \frac{\left|\mathbf{h}_i^H \mathbf{w}_m\right|^2}{\sum_{j \neq m} \left|\mathbf{h}_i^H \mathbf{w}_j\right|^2 + M/\bar{\gamma}_i}. \tag{5.8}$$

M out of N users will be selected in time-slot n according to the procedure shown in Figure 5.2. The user $i_m^*(n)$ that is going to be scheduled on beam m (provided it is not already selected on any other beam) is given by the following scheduling algorithm:

$$i_m^*(n) = \underset{1 \leq i \leq N, i \notin \mathcal{S}}{\operatorname{argmax}} \left(\frac{r_{i,m}(n)}{\alpha_i} \right), \tag{5.9}$$

where $r_{i,m}(n) = W \log_2(1 + \eta_{i,m}(n))$ is the instantaneous rate of user i on beam m in time-slot n, $\mathcal{S}$ is the set of already selected users, and α_i is a constant to be optimized.

To find the α_is, we assume that the PDFs of the users' channel gains are known, and that we have an ideal link adaptation protocol and block-fading. We proceed as in Section 3.3, and again define the random variable $S_i \triangleq \frac{R_i}{\alpha_i}$, where R_i is the random variable describing the rate of user i. For flat, block-fading AWGN channels, the maximal value of S_i for user i within a time-slot (block) with SINR η is then given as

$$S_i(\eta) = \frac{W \log_2(1 + \eta)}{\alpha_i}. \tag{5.10}$$

The PDF for the normalized rate $S_i = s$ for user i can be written as

$$f_{S_i}(s) = \left. \frac{f_{\eta_i}(\eta)}{\frac{dS_i(\eta)}{d\eta}} \right|_{\eta = 2^{\frac{s \cdot \alpha_i}{W}} - 1}, \tag{5.11}$$

where $f_{\eta_i}(\eta)$ is the PDF of the SINR of user i. Assuming Rayleigh fading channel gains, $f_{\eta_i}(\eta)$ is given as [Sharif and Hassibi, 2005; Vicario, Bosisio, Anton-Haro, and Spagnolini, 2008],

$$f_{\eta_i}(\eta) = \frac{e^{-\frac{M\eta}{\bar{\gamma}_i}}}{(1+\eta)^M} \left(\frac{M}{\bar{\gamma}_i}(1+\eta) + M - 1 \right). \tag{5.12}$$

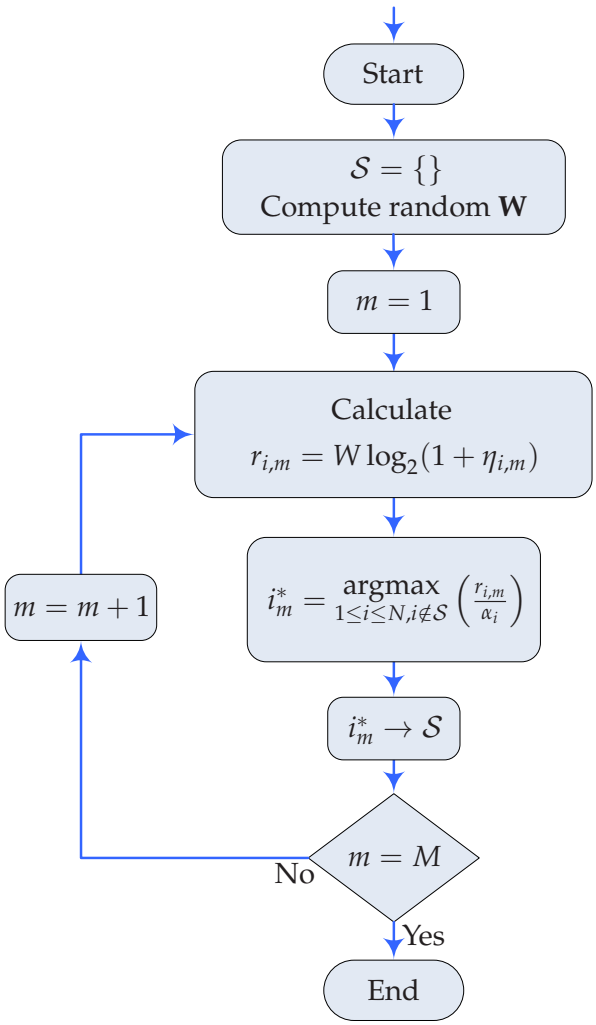


FIGURE 5.2: User selection procedure with random beams for all the users.

(5.11) thus becomes

$$f_{S_i}(s) = \frac{\alpha_i \ln(2)}{W} \frac{e^{-\frac{M(2^{\frac{s \cdot \alpha_i}{W}} - 1)}{\bar{\gamma}_i}}}{(2^{\frac{s \cdot \alpha_i}{W}})^{M-1}} \left(\frac{M}{\bar{\gamma}_i} 2^{\frac{s \cdot \alpha_i}{W}} + M - 1 \right), \tag{5.13}$$

and using simple integration, the corresponding CDF can be expressed as

$$F_{S_i}(s) = \int_0^s f_{S_i}(x)dx = 1 - \frac{e^{-\frac{M(2^{\frac{s\cdot\alpha_i}{W}}-1)}{\bar{\gamma}_i}}}{(2^{\frac{s\cdot\alpha_i}{W}})^{M-1}}. \tag{5.14}$$

Inserting (5.13) and (5.14) in (3.12)-(3.14), and then combining (3.5), (3.12), and (3.14) we again obtain $3N$ equations in $3N$ unknowns, and can thus find the optimal α_is.

5.3.2 Transmit Beamforming for the First Selected User

In this scheme, transmit beamforming is used for the first selected user, whereas random beams are used for the remaining users. The idea behind this scheme is to select the beamforming vectors such that the first selected user does not experience any interference from other selected users. The user selection procedure for time-slot n is shown in Figure 5.3. The first user $i_1^*(n)$ that is going to be scheduled on beam 1 in time-slot n is given by the following scheduling algorithm:

$$i_1^*(n) = \underset{1\leq i\leq N}{\text{argmax}} \left(\frac{r_{i,1}(n)}{\beta_i}\right), \tag{5.15}$$

where $r_{i,1}(n)$ is the instantaneous rate of user i on beam 1 in time-slot n, and is calculated from the SINR of user i assuming transmit matched filtering (and no interference). The SINR in this case is given as

$$\eta_{i,1} = \frac{\bar{\gamma}_i}{M}\left|\mathbf{h}_i^H\mathbf{w}_i\right|^2, \tag{5.16}$$

where

$$\mathbf{w}_i = \frac{\mathbf{h}_i}{\|\mathbf{h}_i\|}. \tag{5.17}$$

The PDF of the SINR of user i in this case (assuming Rayleigh fading) is given as [Hammarwall, Bengtsson, and Ottersten, 2008]

$$f_{\eta_i}(\eta) = \frac{M}{\bar{\gamma}_i}e^{-\frac{M\eta}{\bar{\gamma}_i}}. \tag{5.18}$$

The constant β_i in (5.15) is similar to α_i, and can be calculated in the same way. We again define the random variable $S_i \triangleq \frac{R_i}{\beta_i}$. The maximal value of the metric S_i within a time-slot (block) with SINR η is then given as

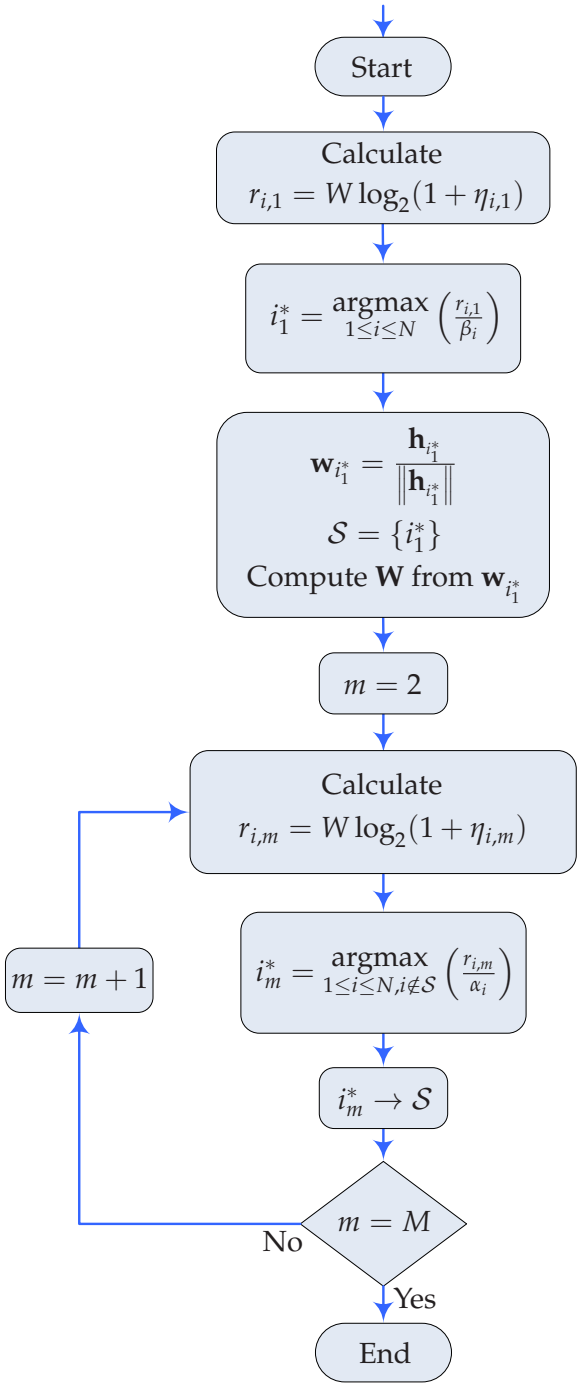


FIGURE 5.3: User selection procedure with transmit beamforming for the first selected user.

$$S_i(\eta) = \frac{W \log_2(1+\eta)}{\beta_i}. \tag{5.19}$$

The PDF $f_{S_i}(s)$ and the CDF $F_{S_i}(s)$ for the normalized rate $S_i = s$ for user i (the first selected user) are given as follows:

$$f_{S_i}(s) = \frac{M\beta_i \ln(2)}{W\bar{\gamma}_i} 2^{\frac{s\cdot\beta_i}{W}} e^{-\frac{M(2^{\frac{s\cdot\beta_i}{W}}-1)}{\bar{\gamma}_i}}. \tag{5.20}$$

$$F_{S_i}(s) = \int_0^s f_{S_i}(x)dx = 1 - e^{-\frac{M(2^{\frac{s\cdot\beta_i}{W}}-1)}{\bar{\gamma}_i}}. \tag{5.21}$$

The optimal β_is can be obtained using the method described in the previous section by replacing α_i with β_i.

The beamforming vector $\mathbf{w}_{i_1^*}$ of the first selected user serves as a basis vector of the beamforming matrix $\mathbf{W}$ to be used for transmission. The matrix $\mathbf{W}$ is obtained by applying the Gram-Schmidt orthogonalization process [Cohen, 1993].

The procedure for selecting the users on the remaining $M-1$ beams is exactly the same as the one described in the previous section. The user $i_m^*(n)$ that is going to be scheduled on beam m ($= 2, \cdots, M$) in time-slot n is thus given as

$$i_m^*(n) = \underset{1\le i\le N, i\notin\mathcal{S}}{\operatorname{argmax}} \left(\frac{r_{i,m}(n)}{\alpha_i}\right), \tag{5.22}$$

where $r_{i,m}(n)$, the instantaneous rate of user i on beam m in time-slot n is calculated using the SINR $\eta_{i,m}(n)$ given in (5.8), and $\mathcal{S}$ is the set of already selected users. The constant α_i can be obtained from the solution of (3.5), (3.12), and (3.14).

Note that the base station now needs to calculate both α_is and β_is. Hence this scheme is more complex as compared to the first one. However the performance improvement is significant, as will be shown in Section 5.6.

5.4 Improving the Short-Term Performance

Similar to the MTGS algorithm in Chapter 3, the scheduling algorithms in the previous sections are designed for long time-windows. We again need to *adapt* the values of α_i and β_i to the actual resource allocation, in order to improve performance for shorter time-windows. In this chapter, we make

use of the AMTGS2 algorithm given in Section 3.5.2, and modify it a little for the case of multiple beams. The algorithm then works as follows:

For the promised throughput guarantees B_i, select the users $i_1^*(n)$ and $i_m^*(n)$ as follows:

$$i_1^*(n) = \underset{1\leq i\leq N}{\operatorname{argmax}} \left(v_i(n-1)\frac{r_{i,1}(n)}{\beta_i} \right), \tag{5.23}$$

and

$$i_m^*(n) = \underset{1\leq i\leq N, i\notin\mathcal{S}}{\operatorname{argmax}} \left(v_i(n-1)\frac{r_{i,m}(n)}{\alpha_i} \right), \tag{5.24}$$

where $v_i(n)$ is given in (3.22).

5.5 Simulation Parameters

For the simulations in the next section, we consider an LTE network where the time-slot length is 1 ms [Freescale Semiconductor, 2007].

We assume that the length of the time-window is $T_W = 80$ ms. Thus T_W contains $K = 80$ time-slots for the LTE network.

We assume that there are $N = 10$ users in our system, and the user channels are Rayleigh distributed with constant average CNRs that are distributed with an average CNR of 15 dB, as given in Table 2.1.

The number of antennas at the base station is assumed to be $M = 3$, whereas each user has a single antenna.

For simplification, we also assume that identical throughput guarantees are promised to all the users.

Each value in the plots is an average over 500 Monte Carlo simulations.

5.6 Numerical Results

Figure 5.4 shows the TGVP performance in the network where random beamforming is used for all the selected users. We again name our non-adaptive and adaptive algorithms as MTGS and AMTGS2 respectively. We compare our scheduling policies with the maximum SINR scheduling scheme based on random beamforming [Sharif and Hassibi, 2005] and the multi-user PFS scheme defined in [Kountouris and Gesbert, 2005]. As expected, the performance of our non-adaptive algorithm is not very good since it is designed for a long time-window with many time-slots. The AMTGS2

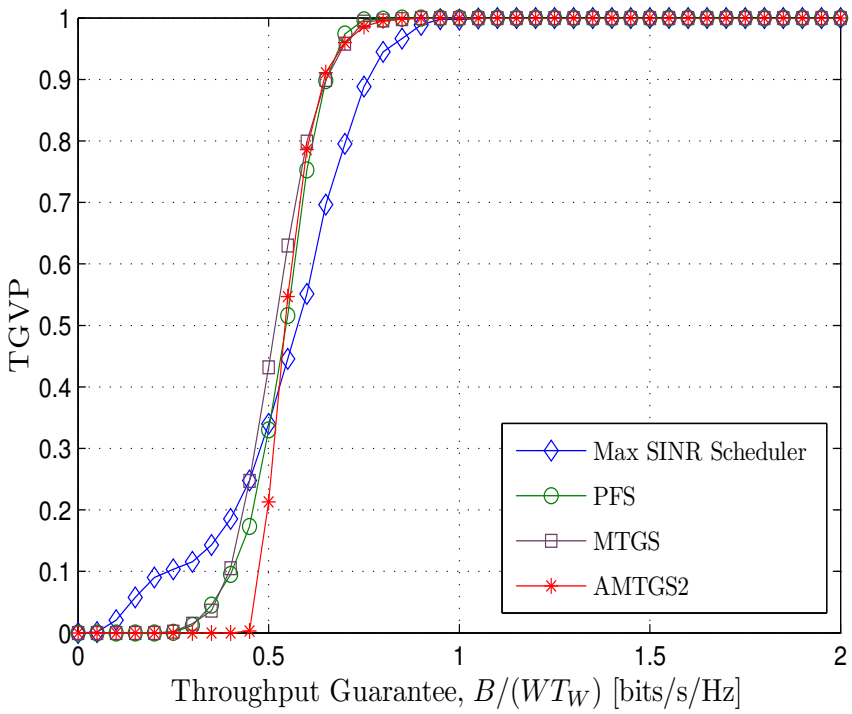


FIGURE 5.4: TGVP for 10 users in an LTE network with random beams used for all selected users. Plotted for a time-window $T_W = 80$ ms that contains 80 time-slots. Each value in the plot is an average over 500 Monte Carlo simulations.

algorithm performs better than the remaining algorithms. However, the performance improvement is not that significant.

Figure 5.5 shows the TGVP performance in the network where transmit beamforming is used for the first selected user, and random beams are used for the remaining users. For a fair comparison, we use the same beamforming strategy for the maximum SINR scheduling and the multi-user PFS schemes. We see that the AMTGS2 algorithm now performs significantly better than all the other scheduling algorithms. The throughput guarantee that can be promised with close to unity probability (TGVP ≈ 0) with this algorithm is almost twice as large as for the PFS algorithm. It is also interesting to observe that the performance of the MTGS algorithm is also better than other algorithms. The results in this case are much better than the one

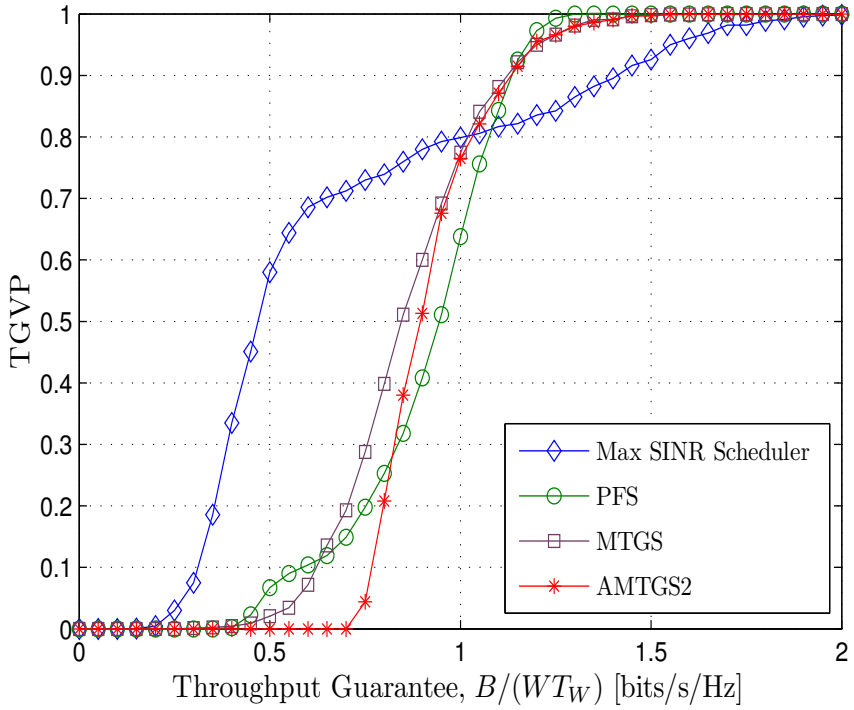


FIGURE 5.5: TGVP for 10 users in an LTE network with transmit beamforming for the first selected user. Plotted for a time-window $T_W = 80$ ms that contains 80 time-slots. Each value in the plot is an average over 500 Monte Carlo simulations.

with random beams for all the users, because now the first user does not experience any interference from other users.

5.7 Summary and Discussion

In this chapter, we have extended the work in Chapter 3 to the case of MIMO broadcast channels. We have proposed scheduling algorithms that make use of orthonormal random beamforming and try to fulfill the throughput guarantees promised to all the users in a network. The scheduling algorithms are designed for two different beamforming scenarios. In the first scenario, random beams are used for all the selected users, whereas transmit beamforming is used for the first selected user in the second sce-

nario. To further improve the short-term performance, we also modified the AMTGS2 algorithm given in Chapter 3 for multiple beams. Results from our simulations show that the adaptive scheduling algorithm with transmit beamforming for the first selected user performs significantly better than other well-known scheduling algorithms, since the first user does not see any interference from other users in this case. We can thus conclude that if we reduce the interference experienced by all the users, the performance may be improved even further.

Chapter 6

Joint Power and Bandwidth Allocation for Discrete-Rate Link Adaptation

In the previous chapters, we have focused on various aspects of the throughput guarantees provided to all the users in wireless networks. Now we shift our focus to the second half of our motivation, which is to design the ACM scheme for multi-user scenario, and to analyze the impact of imperfect channel knowledge on the designed ACM scheme.

In practical wireless networks, the available transmission power and bandwidth of the network are limited. Equal allocation of bandwidth and power to all users may not be efficient for multi-user networks. Therefore, joint bandwidth and power allocation for wireless multi-user networks is essential in order to improve the network performance. In literature [Gong, Vorobyov, and Tellambura, 2011; Julian, Chiang, O'Neill, and Boyd, 2002; Kumaran and Viswanathan, 2005; Phan, Le-Ngoc, Vorobyov, and Tellambura, 2009], joint bandwidth and power allocation has been studied for i) maximizing the sum rate of all users, ii) maximizing the rate of the worst user and iii) minimizing the average power consumption of all users. However, most of the research has focused on continuous rate, power, and bandwidth allocations, i.e, it is assumed that there are infinite number of rates, power and bandwidth levels available. However, this is not the case with practical systems. In this chapter, we therefore consider the issue of joint discrete power and bandwidth allocation for discrete-rate multi-user link

adaptation.

Link adaptation or ACM has been shown to increase throughput in wireless communication systems affected by fading [Gjendemsjø et al., 2008; Goldsmith and Chua, 1997; Holm et al., 2003]. In [Holm et al., 2003], single-user ACM systems are designed such that, given the number of codes C and capacity-approaching codes for AWGN channels, the chosen CNR thresholds and corresponding rates (codes) are optimal with respect to maximal average spectral efficiency. In [Gjendemsjø et al., 2008], continuous and discrete power adaptation for discrete-rate single-user link adaptation with perfect channel knowledge is investigated. The resulting transmission schemes enable very high transmission rates using a limited number of codes. In practice, the channel prediction is not perfect. However, the system can be made more robust towards channel prediction errors by increasing the thresholds in order to be more conservative in the choice of codes [Jetlund et al., 2004].

In this chapter, we analyze the issue of joint power and bandwidth allocation for discrete rates by extending the idea of [Gjendemsjø et al., 2008] to multi-user systems, for example, CDMA systems that allow scheduling of multiple users in each time slot by multiplexing them on different codes (see Section 1.7.3), or OFDM systems in which different carriers are assigned to different users in a single scheduling interval (see Section 1.7.2). To be more specific, we also assume that the channel state information available is imperfect, and then discuss how the ACM scheme can be designed for i) sum rate maximization and ii) average power minimization in a multi-user system.

The remainder of the chapter is organized as follows. The system model and the ACM is introduced in Sections 6.1 and 6.2. In Section 6.3, we discuss the joint power and bandwidth problem for sum rate maximization and average power minimization. Simulation parameters are provided in Section 6.4. We present our numerical results in Section 6.5, and give a brief summary/discussion in Section 6.6.

This chapter is based on [Rasool and Øien, 2012a].

6.1 System Model

Consider a wireless downlink system in which time is divided into slots, and in each time-slot the base station resource, consisting of bandwidth

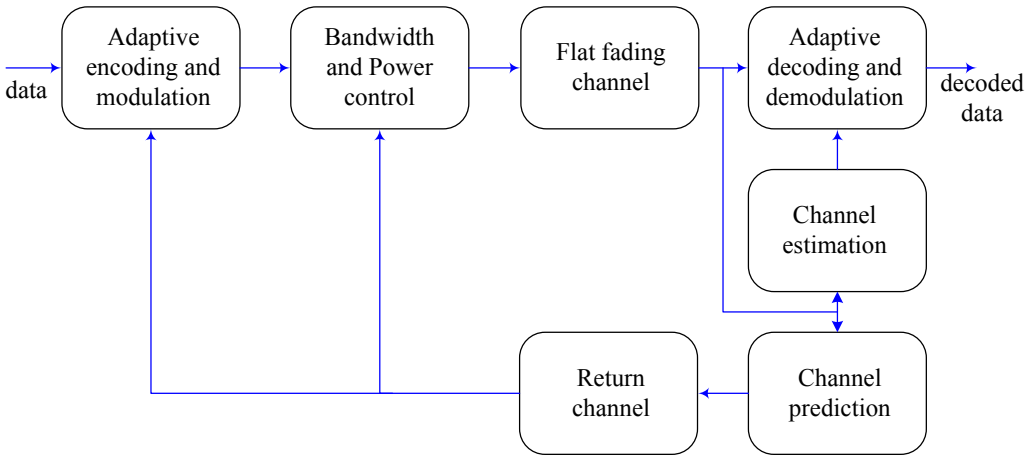


FIGURE 6.1: System Model between the base station and user i.

and power, is divided among N users, based on the channel feedback from them. It is assumed that the total number of users in the system remains constant and equal to N [1]. We further assume that no two users can share the same bandwidth, and thus the users do not interfere with each other.

The wireless system between the base station and user i is shown in Figure 6.1. The corresponding discrete-time channel between the base station and user i is given by (4.2), with the time-varying gain $g_i(n) = |h_i(n)|$ and AWGN noise $n_i(n)$. The fading is assumed to be slow and flat. Let $\bar{P}$ denote the average transmit power without power adaptation, and W [Hz] is the transmission bandwidth allocated to user i without adaptation, such that the total available bandwidth is $W_t = NW$ [Hz] [2]. The instantaneous pre-adaptation CNR of user i at time-slot n is then

$$\gamma_i(n) = \frac{g_i^2(n)\bar{P}}{N_0 W},$$

where N_0 [Watt/Hz] is the noise power spectral density. The average pre-adaptation received CNR of user i is denoted by $\bar{\gamma}_i$. Assuming that the transmit power and bandwidth are adapted instantaneously at time-slot n

[1] Otherwise the system needs to run the optimization problem again for each new value of N.

[2] Equal power and bandwidth allocation is assumed when there is no adaptation.

based on the *predicted* received CNR, $\hat{\gamma}_i(n)$, we denote the transmit power and bandwidth as $P_i(\hat{\gamma}_i(n))$ and $W_i(\hat{\gamma}_i(n))$ respectively. The received post-adaptation CNR of user i at time-slot n is then given by

$$\frac{P_i(\hat{\gamma}_i(n))}{\bar{P}} \frac{W}{W_i(\hat{\gamma}_i(n))} \gamma_i(n).$$

It should be noted that the users will have different noise powers after adaptation, since they will be allocated different bandwidths.

We again assume that a maximum a posteriori (MAP) optimal predictor is employed to predict $\hat{\gamma}_i(n)$ (see Section 4.1.1). In this case the predicted CNR follows an exponential distribution with expectation $\bar{\hat{\gamma}}_i = \rho_i \bar{\gamma}_i$, where ρ_i is the normalized correlation between the actual and predicted pre-adaptation CNRs for user i.

6.1.1 Adaptive Coding and Modulation

We assume that the base station has a set of C codes, with L power levels per code. Following [Gjendemsjø et al., 2008], the range of CNRs is thus divided into $CL + 1$ intervals, which are defined by the *CNR thresholds* $\gamma_{c,l}$ for $c = 1, \cdots, C$ and $l = 1, \cdots, L$, as illustrated in Figure 6.2. We let $\gamma_{0,1} = 0$ and $\gamma_{C+1,1} = \infty$.

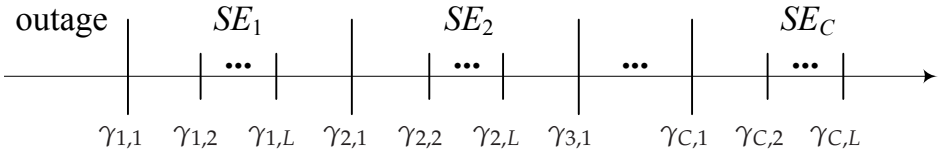


FIGURE 6.2: The range of γ is divided into intervals, where $\gamma_{c,l}$ are the CNR thresholds.

When γ_i [3] is in the interval $[\gamma_{c,1}, \gamma_{c+1,1})$, we select code c with spectral efficiency SE_c. Within this interval the transmission rate is constant, but the transmitted power and bandwidth can be adapted in order to achieve the system objective.

[3] The time reference n is omitted from now on for simplicity.

6.1.2 Code Selection

Since the actual CNR γ_i is unknown at the base station, a new set of thresholds are defined that are used to select a code based on the predicted CNR $\hat{\gamma}_i$ [Jetlund et al., 2004]. The *switching thresholds* are denoted as $\hat{\gamma}_{c,l}$ with $\hat{\gamma}_{0,1} = 0$ and $\hat{\gamma}_{C+1,1} = \infty$. Code c is now selected when $\hat{\gamma}_i$ lies in the interval $[\hat{\gamma}_{c,1}, \hat{\gamma}_{c+1,1})$.

For perfect channel prediction ($\rho_i = 1$), the CNR thresholds in Figure 6.2 will be used as the switching thresholds, i.e. $\hat{\gamma}_{c,l} = \gamma_{c,l}$. However, the normalized correlation ρ_i can be lower due to, for example, lower $\bar{\gamma}_i$, larger prediction delay or higher Doppler frequency shift. This will cause a mismatch between the predicted and actual CNR, resulting in an increased BER, since the actual CNR may sometimes fall into a lower indexed interval than the predicted CNR [Jetlund et al., 2004]. Similar to our analysis in Chapter 4, this cannot be avoided completely, but we can control the probability of this event by demanding [Jetlund et al., 2004]

$$\Pr[\gamma_i < \gamma_{c,l} | \hat{\gamma}_i \geq \hat{\gamma}_{c,l}] \leq \epsilon, \tag{6.1}$$

where ϵ is a small constant chosen by the system designer. The probability of the complementary event is again given as [Jetlund et al., 2004; Øien et al., 2002]

$$\begin{aligned} 1 - \epsilon &= \Pr[\gamma_i > \gamma_{c,l} | \hat{\gamma}_i = \hat{\gamma}_{c,l}] \\ &= Q\left(\frac{\sqrt{\hat{\gamma}_{c,l}}}{\sqrt{\bar{\gamma}_i(1-\rho_i)/2}}, \frac{\sqrt{\gamma_{c,l}}}{\sqrt{\bar{\gamma}_i(1-\rho_i)/2}} \right), \end{aligned} \tag{6.2}$$

where $Q(a, b)$ is the Marcum-Q function given in 4.15.

6.2 Average Rate for ACM

It is assumed that the fading is slow enough for capacity-achieving codes for AWGN channels to be employed, giving relatively tight upper bounds on the average rate [Gjendemsjø et al., 2008]. The C different codes must be selected such that they achieve capacity at the lower end of the corresponding CNR intervals, in order to have arbitrarily small BER throughout the whole of the interval. Assuming that capacity-achieving codes designed

for AWGN channels are used, the spectral efficiency of code c is given as

$$SE_c = \log_2\left(1 + \frac{P_i(\hat{\gamma}_{c,1})}{\bar{P}} \frac{W}{W_i(\hat{\gamma}_{c,1})} \gamma_{c,1}\right) \text{ [bits/s/Hz]},$$

and the average rate $\bar{R}_i$ of user i for the ACM scheme is given as

$$\bar{R}_i = \sum_{c=1}^{C} W_i(\hat{\gamma}_{c,1}) \cdot SE_c \cdot \mathcal{P}_{c,i} \text{ [bits/s]}, \tag{6.3}$$

where $\mathcal{P}_{c,i}$ is the probability of selecting code c for user i, and is given as

$$\mathcal{P}_{c,i} = \int_{\hat{\gamma}_{c,1}}^{\hat{\gamma}_{c+1,1}} f_{\hat{\gamma}_i}(\hat{\gamma}) d\hat{\gamma}, \tag{6.4}$$

where $f_{\hat{\gamma}_i}(\hat{\gamma})$ is the PDF of $\hat{\gamma}_i$ given in (4.11).

Similar to the derivation of the back-off CNR $\gamma_b(\epsilon)$ in Section 4.2, the CNR thresholds $\gamma_{c,l}(\epsilon)$ can be computed as

$$\gamma_{c,l}(\epsilon) = \begin{cases} \left(\sqrt{\frac{\bar{\gamma}_i(1-\rho_i)}{2}} \cdot q_b\left(\frac{\sqrt{\hat{\gamma}_{c,l}}}{\sqrt{\frac{\bar{\gamma}_i(1-\rho_i)}{2}}}, \epsilon\right)\right)^2, & \rho_i < 1 \\ \hat{\gamma}_{c,l}, & \rho_i = 1 \end{cases} \tag{6.5}$$

The average rate $\bar{R}_i$ in (6.3) can now be written as

$$\bar{R}_i(\epsilon) = \sum_{c=1}^{C} W_i(\hat{\gamma}_{c,1}) \log_2\left(1 + \frac{P_i(\hat{\gamma}_{c,1})}{\bar{P}} \frac{W}{W_i(\hat{\gamma}_{c,1})} \gamma_{c,1}(\epsilon)\right) \int_{\hat{\gamma}_{c,1}}^{\hat{\gamma}_{c+1,1}} f_{\hat{\gamma}_i}(\hat{\gamma}) d\hat{\gamma}. \tag{6.6}$$

6.3 Joint Bandwidth and Power Adaptation for ACM

Different optimization objectives can be considered while allocating resources in wireless multi-user networks. We shall consider i) sum rate maximization and ii) average power minimization in our work. We now formulate and solve the problem of bandwidth and power allocation for each of these two different objectives.

6.3.1 Sum Rate Maximization

In applications without delay constraints, a high data rate from any user in the network is favorable. Thus, it is desirable to allocate the resources to

maximize the overall network performance, e.g., the sum rate of all users. The joint bandwidth and power allocation problem aiming at maximizing the sum rate for the network can be formulated as

$$\bar{R}_{max} = \max \sum_{i=1}^{N} \bar{R}_i(\epsilon), \tag{6.7}$$

subject to the average power constraint

$$\sum_{c=0}^{C} \int_{\hat{\gamma}_{c,1}}^{\hat{\gamma}_{c+1,1}} P_i(\hat{\gamma}) f_{\hat{\gamma}_i}(\hat{\gamma}) d\hat{\gamma} \leq \bar{P}, \tag{6.8}$$

and the total bandwidth constraint

$$\sum_{i=1}^{N} W_i(\hat{\gamma}_i) \leq W_t. \tag{6.9}$$

We now consider different adaptation strategies for allocating power and bandwidth.

6.3.1.1 Constant Bandwidth Constant Power Adaptation

Under this strategy, the transmission bandwidth allocated to user i is now restricted to be constant in the interval $[\hat{\gamma}_{1,1}, \hat{\gamma}_{C+1,1})$. I.e.,

$$\frac{W_i(\hat{\gamma}_i)}{W} = \begin{cases} \omega_i \hat{\gamma}_{1,1}, & if\ \hat{\gamma}_{c,1} \leq \hat{\gamma}_i \leq \hat{\gamma}_{c+1,1}, \\ & 1 \leq c \leq C \\ 0, & if\ \hat{\gamma}_i < \hat{\gamma}_{1,1} \end{cases}, \tag{6.10}$$

for all $i = 1, \cdots, N$. The constraint in (6.9) then becomes

$$\sum_{i=1}^{N} \omega_i \hat{\gamma}_{1,1} = N, \text{ for all } c = 1, \ldots, C. \tag{6.11}$$

Furthermore, a single transmission power is now used for all codes for user i. From (6.8) the optimal constant power policy is [Gjendemsjø et al., 2008]

$$\frac{P_i(\hat{\gamma}_i)}{\bar{P}} = \begin{cases} \frac{1}{1-F_{\hat{\gamma}_i}(\hat{\gamma}_{1,1})}, & if\ \hat{\gamma}_{c,1} \leq \hat{\gamma}_i \leq \hat{\gamma}_{c+1,1}, \\ & 1 \leq c \leq C \\ 0, & if\ \hat{\gamma}_i < \hat{\gamma}_{1,1} \end{cases}, \tag{6.12}$$

where $F_{\hat{\gamma}_i}(\cdot)$ denotes the CDF of $\hat{\gamma}_i$. We obtain $\bar{R}_{max,C}$:

$$\bar{R}_{max,C} = \max \sum_{i=1}^{N} \sum_{c=1}^{C} \omega_i \hat{\gamma}_{1,1} \log_2 \left(1 + \frac{\gamma_{c,1}(\epsilon)}{\omega_i \hat{\gamma}_{1,1} \left(1 - F_{\hat{\gamma}_i}(\hat{\gamma}_{1,1})\right)}\right) \int_{\hat{\gamma}_{c,1}}^{\hat{\gamma}_{c+1,1}} f_{\hat{\gamma}_i}(\hat{\gamma}) d\hat{\gamma}, \tag{6.13}$$

subject to constraint (6.11). The variables $\{\omega_i\}_{i=1}^{N}$ and $\{\hat{\gamma}_{c,1}\}_{c=1}^{C}$ are then found by numerical optimization. This can, for example, be achieved in MATLAB by using the function `fmincon`.

6.3.1.2 Constant Bandwidth Discrete Power Adaptation

The transmission bandwidth allocated to user i is again restricted to be constant in the interval $[\hat{\gamma}_{1,1}, \hat{\gamma}_{C+1,1})$, given by (6.10). However, there are now $L \geq 1$ power levels within each of the C intervals [Gjendemsjø et al., 2008]. For each interval $[\hat{\gamma}_{c,1}, \hat{\gamma}_{c+1,1})$ we again use a capacity-achieving code which ensures an arbitrarily low probability of error for any AWGN channel with received CNR greater than or equal to $\frac{P_i(\hat{\gamma}_{c,1})}{\bar{P}} \frac{W}{W_i(\hat{\gamma}_{c,1})} \gamma_{c,1}(\epsilon)$, imposing the following restriction (similar to the one in [Gjendemsjø et al., 2008]):

$$\frac{P_i(\hat{\gamma}_i)}{\bar{P}} \frac{W}{W_i(\hat{\gamma}_i)} \gamma_i(\epsilon) \geq \frac{P_i(\hat{\gamma}_{c,1})}{\bar{P}} \frac{W}{W_i(\hat{\gamma}_{c,1})} \gamma_{c,1}(\epsilon). \tag{6.14}$$

Since the rate and the bandwidth is restricted to be constant in each interval, the transmitted power can be reduced when the channel conditions are more favorable. We define

$$\beta_c = \frac{P_i(\hat{\gamma}_{c,1})}{\bar{P}} \frac{W}{W_i(\hat{\gamma}_{c,1})} \gamma_{c,1}(\epsilon).$$

The value of β_c thus corresponds to the minimum received CNR within the interval c for $1 \leq c \leq C$. The jointly optimal power and bandwidth adaptation scheme is then of the form (following [Gjendemsjø et al., 2008])

$$\frac{P_i(\hat{\gamma}_i)}{\bar{P}} \frac{W}{W_i(\hat{\gamma}_i)} = \begin{cases} \frac{\beta_c}{\gamma_{c,l}(\epsilon)}, & if\ \hat{\gamma}_{c,l} \leq \hat{\gamma}_i \leq \hat{\gamma}_{c+1,l}, \\ & 1 \leq c \leq C, 1 \leq l \leq L \\ 0, & if\ \hat{\gamma}_i < \hat{\gamma}_{1,1} \end{cases}$$

Or

$$\frac{P_i(\hat{\gamma}_i)}{\bar{P}} = \begin{cases} \frac{\omega_i \beta_c \hat{\gamma}_{1,1}}{\gamma_{c,l}(\epsilon)}, & if\ \hat{\gamma}_{c,l} \leq \hat{\gamma}_i \leq \hat{\gamma}_{c+1,l}, \\ & 1 \leq c \leq C, 1 \leq l \leq L \\ 0, & if\ \hat{\gamma}_i < \hat{\gamma}_{1,1} \end{cases}. \tag{6.15}$$

We thus have the following optimization problem with respect to the variables $\{\omega_i\}_{i=1}^{N}$, $\{\beta_c\}_{c=1}^{C}$ and $\{\hat{\gamma}_{c,l}\}_{c=1,l=1}^{C,L}$: Maximize

$$\bar{R}_{max,C\times L} = \sum_{i=1}^{N}\sum_{c=1}^{C} \omega_i \hat{\gamma}_{1,1} \log_2(1+\beta_c) \int_{\hat{\gamma}_{c,1}}^{\hat{\gamma}_{c+1,1}} f_{\hat{\gamma}_i}(\hat{\gamma}) d\hat{\gamma}, \tag{6.16}$$

such that for all $i = 1, \ldots, N$,

$$\sum_{c=1}^{C} \omega_i \beta_c \hat{\gamma}_{1,1} \sum_{l=1}^{L} \frac{1}{\gamma_{c,l}(\epsilon)} \int_{\hat{\gamma}_{c,l}}^{\hat{\gamma}_{c,l+1}} f_{\hat{\gamma}_i}(\hat{\gamma}) d\hat{\gamma} = 1, \tag{6.17}$$

and for all $c = 1, \ldots, C$,

$$\sum_{i=1}^{N} \omega_i \hat{\gamma}_{1,1} = N. \tag{6.18}$$

This problem can again be solved in MATLAB by using the function `fmincon`.

6.3.2 Average Power Minimization

Another relevant design objective (e.g., to prolong battery lifetime and/or reduce interference to co-existing systems) is the minimization of the average power consumption of all users. This minimization is performed under the constraint that the minimum average rate requirements of all users must be satisfied. The corresponding joint bandwidth and power allocation problem can be written as

$$\min \sum_{i=1}^{N}\sum_{c=0}^{C} \int_{\hat{\gamma}_{c,1}}^{\hat{\gamma}_{c+1,1}} P_i(\hat{\gamma}_i) f_{\hat{\gamma}_i}(\hat{\gamma}) d\hat{\gamma}, \tag{6.19}$$

such that

$$\sum_{i=1}^{N} W_i(\hat{\gamma}_i) \leq W_t, \tag{6.20}$$

and

$$\bar{R}_i(\epsilon) \geq r_i. \tag{6.21}$$

Two different adaptation strategies for allocating power and bandwidth are discussed below.

6.3.2.1 Constant Bandwidth Constant Power Adaptation

For constant bandwidth constant power adaptation, the problem in (6.19)-(6.21) becomes

$$\min \sum_{i=1}^{N} P_i(\hat{\gamma}_i)\left(1 - F_{\hat{\gamma}_i}(\hat{\gamma}_{1,1})\right), \tag{6.22}$$

such that for all $c = 1, \ldots, C$,

$$\sum_{i=1}^{N} \omega_i \hat{\gamma}_{1,1} = N, \tag{6.23}$$

and for all $i = 1, \ldots, N$,

$$\sum_{c=1}^{C} \omega_i \hat{\gamma}_{1,1} \log_2\left(1 + \frac{P_i(\hat{\gamma}_i)\gamma_{c,1}(\epsilon)}{\bar{P}\omega_i \hat{\gamma}_{1,1}}\right) \int_{\hat{\gamma}_{c,1}}^{\hat{\gamma}_{c+1,1}} f_{\hat{\gamma}_i}(\hat{\gamma}) d\hat{\gamma} \geq r_i. \tag{6.24}$$

The variables $\{\omega_i\}_{i=1}^{N}$ and $\{\hat{\gamma}_{c,1}\}_{c=1}^{C}$ are then found by using the function `fmincon` in MATLAB.

6.3.2.2 Constant Bandwidth Discrete Power Adaptation

The corresponding optimization problem for constant bandwidth discrete power adaptation becomes

$$\min \sum_{i=1}^{N} \sum_{c=1}^{C} \omega_i \beta_c \hat{\gamma}_{1,1} \sum_{l=1}^{L} \frac{1}{\gamma_{c,l}(\epsilon)} \int_{\hat{\gamma}_{c,l}}^{\hat{\gamma}_{c,l+1}} f_{\hat{\gamma}_i}(\hat{\gamma}) d\hat{\gamma}, \tag{6.25}$$

such that for all $c = 1, \ldots, C$,

$$\sum_{i=1}^{N} \omega_i \hat{\gamma}_{1,1} = N, \tag{6.26}$$

and for all $i = 1, \ldots, N$,

$$\sum_{c=1}^{C} \omega_i \hat{\gamma}_{1,1} \log_2(1 + \beta_c) \int_{\hat{\gamma}_{c,1}}^{\hat{\gamma}_{c+1,1}} f_{\hat{\gamma}_i}(\hat{\gamma}) d\hat{\gamma} \geq r_i. \tag{6.27}$$

The variables $\{\omega_i\}_{i=1}^{N}$, $\{\beta_c\}_{c=1}^{C}$ and $\{\hat{\gamma}_{c,l}\}_{c=1,l=1}^{C,L}$ can then be found by using the function `fmincon` in MATLAB.

The ACM scheme for multi-user scenario can now be designed as follows: For the given system objective, the number of users N, the average

CNRs $\bar{\gamma}_i$ for all the users, the number of codes C, and a bandwidth/power adaptation scheme in mind, find the set of switching thresholds, corresponding CNR thresholds, the bandwidth allocations, and the maximal spectral efficiencies SE_cs. Then design optimal codes for these spectral efficiencies for each N and $\bar{\gamma}_i$s of interest.

6.4 Simulation Parameters

We consider a system with 5 users. All the user channels are Rayleigh distributed with constant average CNRs that are distributed with an average CNR $\bar{\gamma}_t$, and the user with the worst channel has an average CNR of $\frac{3}{4}\bar{\gamma}_t$. For example, for $\bar{\gamma}_t = 10$ dB, the users with the worst and the best channels have average CNRs of 7.5 dB and 11.58 dB respectively, as given in Table 6.1. The rest of the system parameters are chosen as follows:

- carrier frequency $f_c = 1$ GHz,
- user speed $v = 30$ m/s, and
- $\epsilon = 2 \times 10^{-3}$.
- Set 1
  - pilot spacing $\mathcal{L} = 15$,
  - filter length $\mathcal{K} = 200$,
  - prediction lag $\tau = \Delta T_s = 250\mu$s,
- Set 2
  - pilot spacing $\mathcal{L} = 10$,
  - filter length $\mathcal{K} = 250$,
  - prediction lag $\tau = \Delta T_s = 100\mu$s,

We shall use both the sets in our simulations to analyze the effect of imperfect channel prediction. The Set 2 results in higher normalized correlation for all the users. For example, the values of ρ_i for 5 users with $\bar{\gamma}_t = 10$ dB for the two sets are given in Table 6.1.

user i	1	2	3	4	5
CNR $\bar{\gamma}_i$ [dB]	7.50	8.93	10.00	10.86	11.58
ρ_i for Set 1	0.939	0.952	0.961	0.966	0.971
ρ_i for Set 2	0.977	0.983	0.986	0.988	0.990

TABLE 6.1: Values of the average CNRs and the normalized correlation for 5 Rayleigh distributed users, with average CNRs distributed with an average CNR $\bar{\gamma}_t = 10$ dB.

6.5 Numerical Results

6.5.1 Sum Rate Maximization

In this section, we analyze the effect of imperfect channel knowledge on the sum rate of the system. Figure 6.3 shows the set of optimal CNR switching thresholds for $\bar{R}_{max,2\times2}$ as a function of $\bar{\gamma}_t$. We observe that the thresholds for the imperfect case (shown here for Set 2) are higher than the thresholds for $\rho = 1$ (perfect prediction). Since the outage probability depends on $\hat{\gamma}_{1,1}$, imperfect prediction will result in a higher outage probability. Therefore an outage probability constraint can be introduced if an application under consideration has strict real-time or low-delay requirements.

The corresponding minimum received CNR values β_c are depicted in Figure 6.4. For imperfect prediction (Set 1 and Set 2), the received CNR values are lower and therefore the values of designed spectral efficiencies SE_c will also be lower, causing a reduction in the sum rate as will be shown in the next figures.

Under the average power and bandwidth constraints and with perfect channel knowledge, the average sum rates corresponding to $\bar{R}_{max,C\times L}$ and $\bar{R}_{max,C}$ schemes are plotted in Figure 6.5. We see (as noted before for the single-user case in [Gjendemsjø et al., 2008]) that the discrete rate schemes approach the performance of the continuous rate scheme using only a few codes. The effect of imperfect channel knowledge is shown in Figure 6.6 for the $\bar{R}_{max,2\times2}$ scheme. We observe (as expected) that the sum rate is reduced due to imperfect channel prediction. To approach the sum rate of the perfect case, ϵ should be as large as possible. But then we would increase

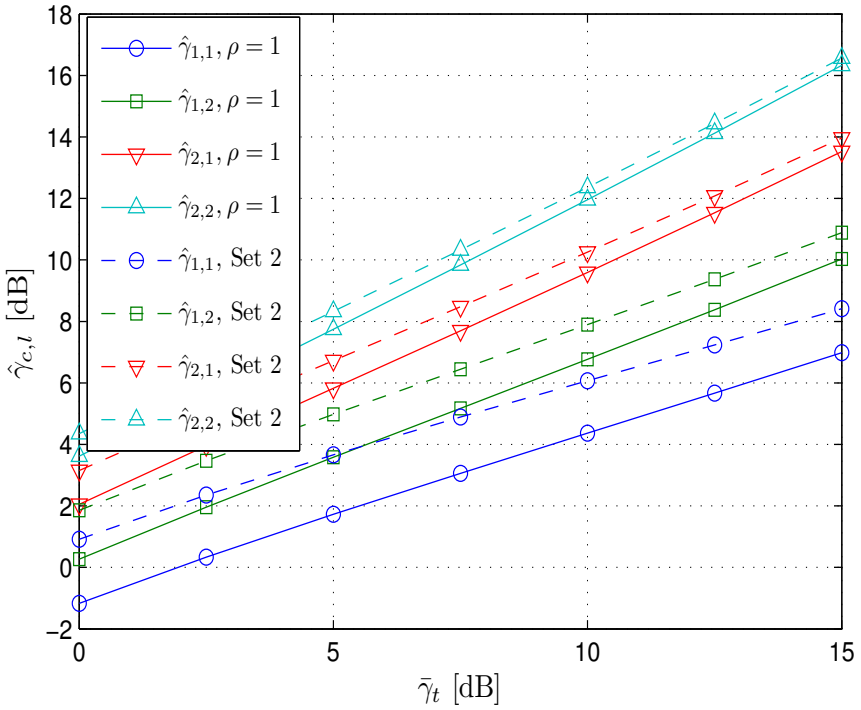


FIGURE 6.3: Switching thresholds for $\bar{R}_{max,2\times2}$ scheme as a function of $\bar{\gamma}_t$ [dB].

the probability that the actual CNR may fall into a lower indexed interval than the predicted CNR, resulting in increased BER. Finding an optimal value of ϵ remains a topic for future research. We have chosen a very small value for ϵ in this section, which results in reduced spectral efficiencies and hence the reduced sum rate. Insertion of equally spaced pilot-symbols also reduces the sum rate by a factor of $1-(\mathcal{L}-1)/\mathcal{L}$. The sum rate for the Set 1 (compared to the Set 2) suffers more due to a combination of small ϵ and higher mismatch between the actual and predicted CNRs.

An example of the optimized power and bandwidth allocation for the $\bar{R}_{max,2\times3}$ scheme is shown in Figure 6.7, for $\bar{\gamma}_t = 10$ dB. The plot is shown for user 1 only, whose average CNR is 7.5 dB. The figure also shows the spectral efficiencies (SE_c) in the two intervals. This figure can be interpreted as follows: The base station will first determine the interval in which the

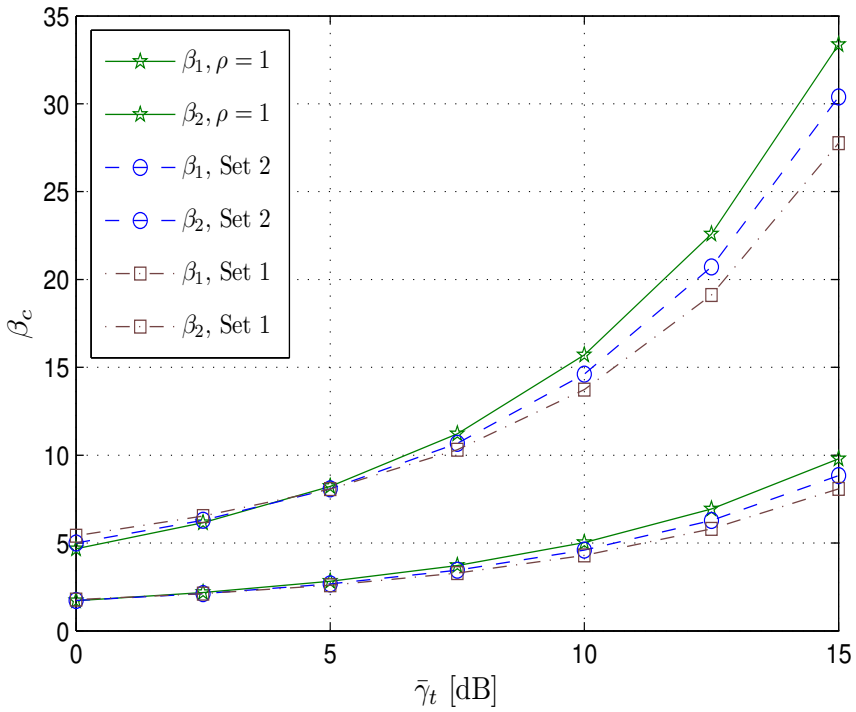


FIGURE 6.4: Minimum received CNR values β_c for $\bar{R}_{max,2\times2}$ scheme as a function of $\bar{\gamma}_t$ [dB].

instantaneous pre-adaptation CNR of user 1, given by $\hat{\gamma}_1$, exists. It will then allocate the corresponding power and bandwidth to the user 1, and send the data at the given rate for that interval using the corresponding SE_c. Within the interval, the rate and the bandwidth is constant, therefore the power is reduced as shown in the figure.

6.5.2 Average Power Minimization

The effect of imperfect channel knowledge on the average power consumed in the system is analyzed in this section. We have not shown the switching thresholds and minimum received CNR curves in this section because they are similar to the ones in the previous section. Figure 6.8 shows average power consumed for constant bandwidth constant power ($\bar{P}_{min,C}$) and constant bandwidth discrete power ($\bar{P}_{min,C\times L}$) schemes using joint power and

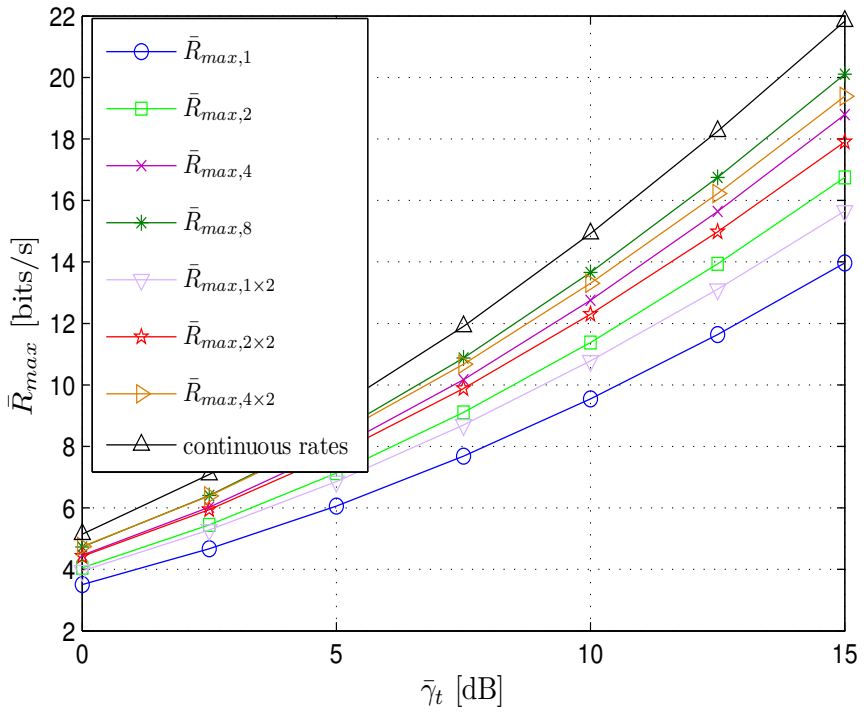


FIGURE 6.5: Various $\bar{R}_{max,C}$ and $\bar{R}_{max,C\times L}$ schemes as a function of $\bar{\gamma}_t$ [dB].

bandwidth optimization, where the base station has perfect channel knowledge. For simplification, we have assumed identical rate requirements for all the users. The required rate is set to $r_i = 1$ bit/s. We can again observe from the figure that the discrete rate schemes approach the performance of the continuous rate scheme using only a few codes. The effect of imperfect channel knowledge is shown in Figure 6.9 for the $\bar{P}_{min,2\times2}$ scheme. We can say that due to imperfect channel prediction, the base station needs to consume more power to fulfill the rate requirements of all the users in the system. The same factors, that were responsible for the reduced sum rate in the previous section, result in increased average power consumption for imperfect channel knowledge. We also observe that the average power consumed for the Set 1 is higher than that for the Set 2. This is again due to the combination of small ϵ and higher mismatch between the actual

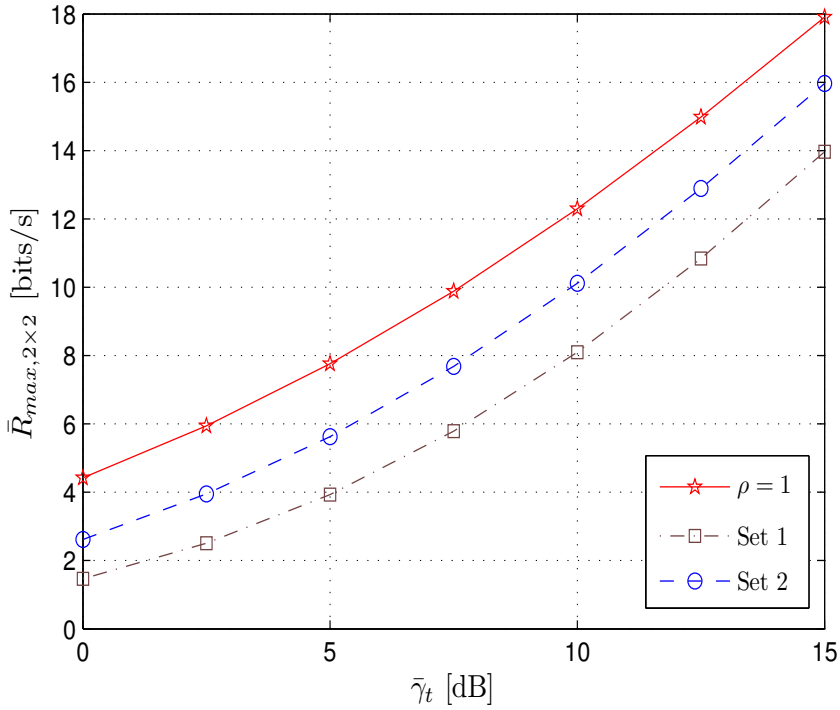


FIGURE 6.6: Effect of imperfect channel prediction on $\bar{R}_{max,2\times2}$ scheme.

and predicted CNRs.

6.6 Summary and Discussion

In this chapter, we have considered the issue of discrete power and bandwidth allocation for discrete-rate multi-user link adaptation with imperfect channel state information. We have shown that with only a few rates (codes), and joint power and bandwidth optimization, the system can approach the performance of a theoretically optimal continuous rate system. We have also observed that the correlation between predicted and actual values of the fading envelope affects the system in the sense that the sum rate is reduced and average power is increased as the normalized correlation is decreased. The outage probability also increases. Therefore for applications with low delay requirements, a constraint on the outage prob-

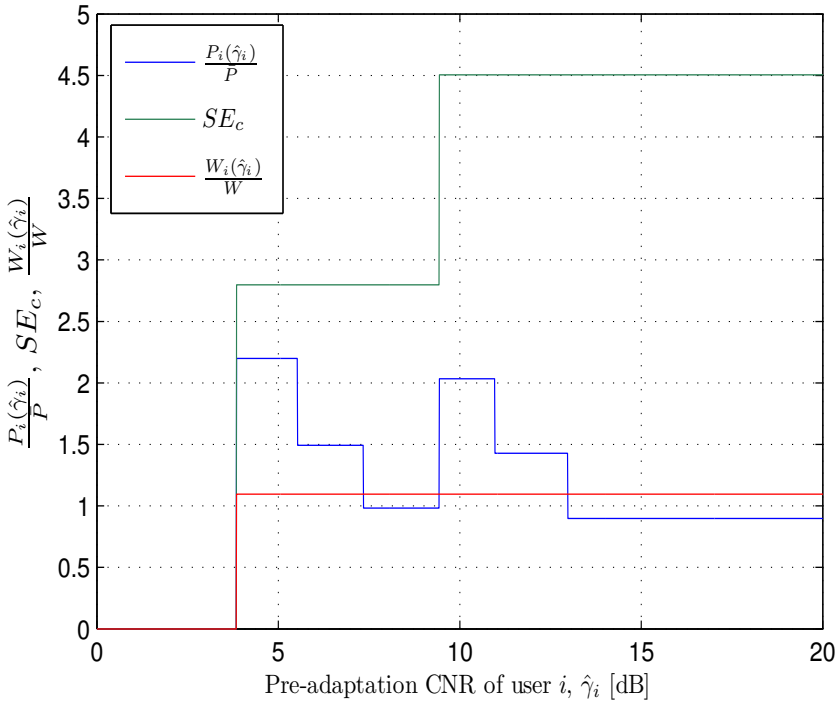


FIGURE 6.7: Bandwidth and power allocation of user 1 for the $\bar{R}_{max,2\times3}$ scheme, whose pre-adaptation average CNR is $\bar{\gamma}_1 = 7.5$ dB, for $\bar{\gamma}_t = 10$ dB. Spectral efficiencies (SE_c) in the two intervals are also shown.

ability may also be applied. Furthermore, since there was no peak power constraint, the system uses higher instantaneous power for the case of imperfect channel state at the base station.

We have only considered constant bandwidth allocation in our work. The reason for this can be explained by the following example. We consider a system with two users ($N = 2$) and two codes ($C = 2$), and suppose we require discrete bandwidth allocation. The bandwidth allocated to user i in the interval c can be denoted by $W_{i,c}$. Then we have to fulfill the following constraints.

$$W_{1,1} + W_{2,1} = W_t, \tag{6.28}$$

$$W_{1,2} + W_{2,2} = W_t, \tag{6.29}$$

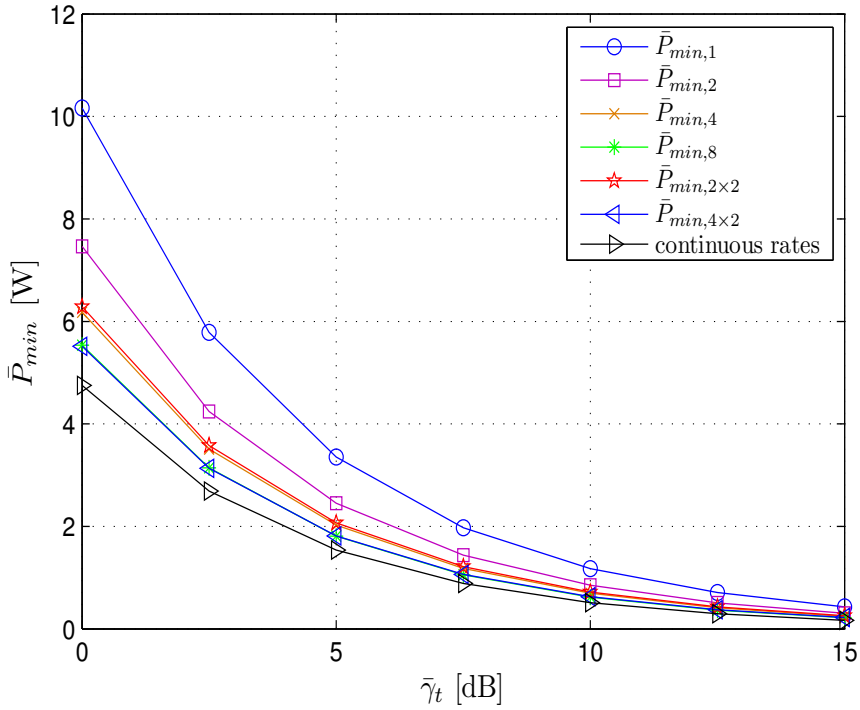


FIGURE 6.8: Various $\bar{P}_{min,C}$ and $\bar{P}_{min,C\times L}$ schemes as a function of $\bar{\gamma}_t$ [dB].

$$W_{1,1} + W_{2,2} = W_t, \tag{6.30}$$

and

$$W_{1,2} + W_{2,1} = W_t. \tag{6.31}$$

From (6.28) and (6.30)

$$W_{2,1} = W_{2,2}, \tag{6.32}$$

and from (6.28) and (6.31)

$$W_{1,1} = W_{1,2}. \tag{6.33}$$

Thus it turns out that the bandwidth allocation is the same (constant) in both the intervals. Therefore, we only need to consider constant bandwidth allocation.

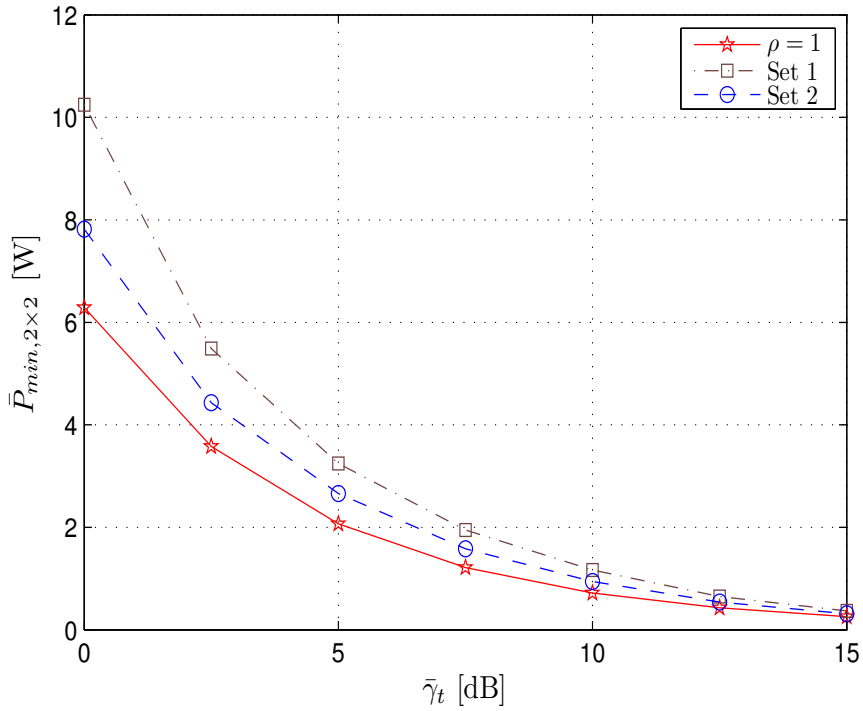


FIGURE 6.9: Effect of imperfect channel prediction on $\bar{P}_{min,2\times2}$ scheme.

Chapter 7

Conclusions

In this dissertation, we have focused on various issues in order to find some answers to the questions raised in Section 1.1. We have proposed an approximate expression for the throughput guarantee violation probability (TGVP), used to analyze the performance of opportunistic scheduling algorithms without conducting experimental investigations. Such an expression can be very useful for the network providers.

We then exploited the channel variations of the user channels to design scheduling algorithms for improved throughput guarantees. Optimization problems have been formulated with an aim at finding optimal scheduling algorithms for maximizing throughput guarantees in a wireless network. We have shown how the solution to such problems can be obtained both when the throughput guarantees are (i) identical and (ii) different for all the mobile users. We have also developed the corresponding adaptive scheduling algorithms, both for the scenarios where single user is scheduled per time-slot and where multiple users are selected in each time-slot (e.g. in MIMO systems). We have considered real-world systems, and analyzed the results for the proposed scheduling schemes. Results from simulations have shown that these algorithms can improve the throughput guarantees in modern cellular networks compared to other well-known scheduling algorithms.

We have further analyzed the performance loss of such algorithms in the case of imperfect channel information, and suggested a rate back-off mechanism to reduce the outage probability in that case. We have also considered the issue of discrete power and bandwidth allocation for discrete-

rate multi-user link adaptation with imperfect channel state information. To be more specific, we have discussed how the ACM scheme can be designed in such a scenario for (i) sum rate maximization and (ii) average power minimization in a multi-user setting.

7.1 Main Contributions of the Work

Chapter 2: We have derived an approximate expression for the TGVP which can be obtained in a time-slotted wireless network with any scheduling policy with a given set of system parameters, known cumulants of the bits transmitted to or from the scheduled user in a time-slot, and a given distribution of the number of time-slots allocated to a user within a time-window. Cramér's theorem from LDT is used to derive this expression. The simulations show that the number of time-slots has a greater effect on the tightness of the approximate TGVP expression than the correlated time-slots. We also observe that the approximate TGVP expression is tighter for longer time-windows. Furthermore, we note that the TGVP expression is more accurate for fast moving users.

Chapter 3: We have formulated an optimization problem which aims at finding the maximum number of bits that can be guaranteed to the users within a time-window for a given set of system parameters. By building on the results in [Borst and Whiting, 2003] and by assuming that the distributions of the users' CNRs are known, we find an optimal scheduling algorithm (MTGS), both for the case where the throughput guarantees are different from user to user, and for the case where the users have the same throughput guarantees. To further improve the short-term performance of this algorithm, we proposed two adaptive and low complexity versions of the MTGS algorithm, i.e. AMTGS1 and AMTGS2. Results from our simulations show that the proposed adaptive algorithms perform significantly better than any of the other well-known scheduling algorithms in Mobile WiMAX-, HSDPA-, LTE-, and WINNER I-based networks. For systems that have many time-slots within the time-window, e.g., for WINNER I, the MTGS algorithm also performs better than all the other well-known algorithms. For a network with heterogeneous

throughput guarantees, the proposed adaptive scheduling algorithms are the only algorithms that support a throughput guarantee close to unity, i.e., TGVP ≈ 0. The AMTGS2 algorithm is simpler in implementation but still provides similar throughput guarantees as provided by the AMTGS1 algorithm. Furthermore, it is also independent of the time-window and therefore overcomes the problem of fixed time-window placement of the AMTGS1 algorithm. The simulations have also shown that the performance of the adaptive algorithms at lower users' speed (strong temporal correlation) also remains the best. We also analyzed the ASSE and fairness of the MTGS algorithm and showed that the throughput fairness of this algorithm is the best.

Chapter 4: The analysis in Chapter 3 assumes that each user estimates or predicts its CNR perfectly, and there is no feedback delay. However, this is not true in practice. Therefore, we have made use of a maximum a posteriori (MAP) channel coefficient predictor to take the delay (prediction lag) and the channel noise into account. We have investigated the effect of imperfect channel prediction and delay on the throughput guarantees promised to all the users in the wireless network, and proposed a rate back-off mechanism to reduce the outage probability of the system. It should be noted that the probability of outage cannot be completely reduced to zero. From the simulations, we observe that the TGVP performance of the system deteriorates when the normalized correlation between the actual and predicted CNR decreases. We also conclude from the performance comparison of Mobile WiMAX with LTE and WINNER I systems that it is better not to predict too far ahead in time when the symbol duration is large. Finally, we also explored the "possibility space" given that a maximum of X% rate loss is acceptable.

Chapter 5: We have extended the work in Chapter 3 by proposing scheduling algorithms that make use of orthonormal random beamforming and try to fulfill the throughput guarantees promised to all the users in a MIMO broadcast channel. The scheduling algorithms are designed for two different beamforming scenarios. In the first scenario, random beams are used for all the selected users, whereas transmit beamforming is used for the first selected user in the second scenario.

To further improve the short-term performance, we also modified the AMTGS2 algorithm given in Chapter 3 for multiple beams. Results from our simulations show that the adaptive scheduling algorithm with transmit beamforming for the first selected user performs significantly better than other well-known scheduling algorithms, since the first user does not see any interference from other users in this case.

Chapter 6: We have considered the issue of discrete power and bandwidth allocation for discrete-rate multi-user link adaptation with imperfect channel state information. We have shown that with only a few rates (codes) and joint power and bandwidth optimization, the system can approach the performance of continuous rate system. We have also observed that the correlation between predicted and actual values of the fading envelope affects the system in the sense that the sum rate is reduced and average power is increased as the normalized correlation is decreased. The outage probability also increases. Therefore for applications with low delay requirements, a constraint on the outage probability may also be applied. Furthermore, since there was no peak power constraint, the system uses higher instantaneous power for the case of imperfect channel state at the transmitter.

7.2 Further Research Directions

In this section we list some topics that can be interesting for further research:

- The approximate TGVP expression derived in Chapter 2 assumes that only one user is scheduled in a time-slot. Extension of this expression for the MIMO or OFDM based systems in which multiple users are selected in a single time-slot is one possible direction for future work.

- Since the approximate TGVP expression assumes uncorrelated time-slots, its accuracy suffers greatly for slow-moving users since for them the time-slots are highly correlated (refer to Chapter 2). Extension of the TGVP expression for correlated time-slots will surely increase the practical value of this expression, and therefore should be investigated.

- We assumed that the number of users are constant within the duration of the time-window over which the throughput guarantees are calculated. This assumption may not always be correct. Analysis of the system performance where this assumption is relaxed is a good topic for further research.
- Call admission control, discussed briefly in Section 1.4.2.2, is often based on the call dropping probability of the system. For networks carrying real-time applications, call admission control schemes in which the call dropping probability is based on the TGVP can be an interesting topic for research.
- Some real-time applications have strict delay constraints which may be violated when the focus is to provide the throughput guarantees to all the users in the network. A study of throughput-delay tradeoff for such applications may be interesting.
- The analysis of Chapter 6 can easily be extended to the scenario of maximization of the minimum average rates.
- OFDM based systems are particularly sensitive to the time and frequency synchronization. Analysis of the impact of synchronization errors on the TGVP performance of OFDM systems is an interesting topic for research.
- In Chapter 5, we have seen great improvement in the TGVP performance by making a single user interference free. Development or inclusion of other beamforming schemes that reduce interference significantly might result in manyfold increase in the system performance, and should be investigated.
- In Chapter 4 and 6, the back-off CNR and the CNR thresholds respectively, are calculated using a pre-defined value of ϵ. How to optimize ϵ for a particular system can be an interesting topic for future research.

Appendix A

An Example to Obtain T_i^N in (2.15)

In this appendix, we show how to obtain the set T_i^N by expanding the product in (2.16). As an example, we take $i = 1$ and $N = 4$. The expansion of the product in (2.16) is given as follows:

$$\begin{aligned}\prod_{\substack{j=1\\ j\neq i}}^{N} F_{\gamma_j}(\gamma) &= \prod_{\substack{j=1\\ j\neq i}}^{N}(1-e^{-\gamma/\bar{\gamma}_j}) = \prod_{\substack{j=1\\ j\neq 1}}^{4}(1-e^{-\gamma/\bar{\gamma}_j}) \\ &= 1-\exp\left(-\frac{\gamma}{\bar{\gamma}_2}\right)-\exp\left(-\frac{\gamma}{\bar{\gamma}_3}\right)-\exp\left(-\frac{\gamma}{\bar{\gamma}_4}\right) \\ &+ \exp\left(-\frac{\gamma}{\bar{\gamma}_2}-\frac{\gamma}{\bar{\gamma}_3}\right)+\exp\left(-\frac{\gamma}{\bar{\gamma}_2}-\frac{\gamma}{\bar{\gamma}_4}\right) \\ &+ \exp\left(-\frac{\gamma}{\bar{\gamma}_3}-\frac{\gamma}{\bar{\gamma}_4}\right)-\exp\left(-\frac{\gamma}{\bar{\gamma}_2}-\frac{\gamma}{\bar{\gamma}_3}-\frac{\gamma}{\bar{\gamma}_4}\right). \quad \text{(A.1)}\end{aligned}$$

The elements of the set T_i^N are found as the terms resulting from taking $(-1/\gamma)$ times the natural logarithm of the absolute value of each term arising from this expansion. T_1^4 is thus given as

$$\begin{aligned}T_1^4 &= \left\{0, \frac{1}{\bar{\gamma}_2}, \frac{1}{\bar{\gamma}_3}, \frac{1}{\bar{\gamma}_4}, \left(\frac{1}{\bar{\gamma}_2}+\frac{1}{\bar{\gamma}_3}\right), \left(\frac{1}{\bar{\gamma}_2}+\frac{1}{\bar{\gamma}_4}\right), \left(\frac{1}{\bar{\gamma}_3}+\frac{1}{\bar{\gamma}_4}\right),\right. \\ &\quad \left.\left(\frac{1}{\bar{\gamma}_2}+\frac{1}{\bar{\gamma}_3}+\frac{1}{\bar{\gamma}_4}\right)\right\}. \quad \text{(A.2)}\end{aligned}$$

The corresponding sign of each term in (A.1) gives sign(τ). We thus

have

$$\text{sign}(\tau) = \{+1, -1, -1, -1, +1, +1, +1, -1\}. \tag{A.3}$$

Appendix B

Derivation of $\Psi(x)$ in (2.21)

In this appendix, we derive $\Psi(x)$ used in (2.21). We begin with the following integral expression:

$$\Psi(x) = \int_0^\infty \ln^2(1+\gamma) e^{-x\gamma} d\gamma, \tag{B.1}$$

where x is a constant.

Changing the variable from γ to $y = \gamma + 1$, (B.1) becomes

$$\Psi(x) = e^x \int_1^\infty \ln^2(y) e^{-xy} dy. \tag{B.2}$$

The integral on the right-hand side of (B.2) can be solved by using [Gradshteyn and Ryzhik, 2007, (4.358.1)]:

$$\begin{aligned} e^{-x}\Psi(x) &= \int_1^\infty \ln^2(y) e^{-xy} dy \\ &= \frac{\partial^2}{\partial a^2} x^{-a}\, \Gamma(a,x)|_{a=1} \\ &= \ln^2(x)\frac{1}{x}\Gamma(1,x) - 2\ln(x)\frac{1}{x}\frac{\partial}{\partial a}\,\Gamma(a,x)|_{a=1} + \frac{1}{x}\frac{\partial^2}{\partial a^2}\,\Gamma(a,x)|_{a=1}, \end{aligned} \tag{B.3}$$

where $\Gamma(a,x)$ is the *incomplete gamma function* [Eric W. Weisstein, d].

Inserting the first and second derivatives of $\Gamma(a,x)$ from [Wolfram Re-

search Inc., c] and setting $a = 1$, we get

$$\begin{aligned} e^{-x}\Psi(x) &= \frac{1}{x}\Big\{\ln^2(x)\Gamma(1,x) - 2\ln(x)\Gamma(1)^2 x\,{}_2\tilde{F}_2(1,1;2,2;-x) \\ &+ 2\Gamma(1,0,x)\ln^2(x) - 2\ln(x)\Gamma(1)\psi(1) + \Gamma(1,x)\ln^2(x) \\ &+ \Gamma(1)(\psi^2(1) + \psi'(1) - \ln^2(x)) \\ &- 2x\,{}_3F_3(1,1,1;2,2,2;-x) + 2x\ln(x)\,{}_2F_2(1,1;2,2;-x)\Big\}, \end{aligned} \tag{B.4}$$

where

- $\Gamma(a)$ is the *gamma function* [Gradshteyn and Ryzhik, 2007, (8.310.1)],
- $\Gamma(a,b,c)$ is the *generalized incomplete gamma function* [Wolfram Research Inc., a],
- $\psi(y) = \frac{d}{dy}\ln\Gamma(y)$ is the *psi function* [Gradshteyn and Ryzhik, 2007, (8.360.1)], and
- ${}_p\tilde{F}_q(a_1,\cdots,a_p;b_1,\cdots,b_q;\cdot)$ is the *regularized hypergeometric function* [Eric W. Weisstein, e].

Inserting the following into (B.4),

- $\Gamma(1) = 1$ [Gradshteyn and Ryzhik, 2007, (8.338.1)],
- $\Gamma(1,0,x) = \Gamma(1) - \Gamma(1,x)$ [Wolfram Research Inc., b],
- $\psi(1) = -C_e$ [Gradshteyn and Ryzhik, 2007, (8.366.1)], where C_e is known as Euler-Mascheroni constant (or Euler's constant),
- $\psi'(1) = \frac{\pi^2}{6}$ [Gradshteyn and Ryzhik, 2007, (8.366.8)], and
- ${}_2\tilde{F}_2(1,1;2,2;-x) = {}_2F_2(1,1;2,2;-x)$, since $\Gamma(2) = 1$ [Gradshteyn and Ryzhik, 2007, (8.338.1)],

we get

$$e^{-x}\Psi(x) = \frac{1}{x}\left\{\frac{\pi^2}{6} + C_e^2 + 2C_e\ln(x) + \ln^2(x) - 2x\,{}_3F_3(1,1,1;2,2,2;-x)\right\}. \tag{B.5}$$

A slight manipulation of (B.5) results in (2.21).

Appendix C

Verification of TGVP Expression's Parameters for the MTGS Algorithm

In this appendix, we analyze the correctness of the parameters of the TGVP expression for the MTGS algorithm, derived in Section 3.6.

Figures C.1 and C.2 gives a comparison of the approximate TGVP expression for the MTGS algorithm with the corresponding Monte Carlo simulated TGVPs for Mobile WiMAX and LTE respectively, for identical throughput guarantees. Each value in the simulated TGVP curves is an average over 1000 Monte Carlo simulations. The TGVP approximate expression is based on the assumption that the time-slots are uncorrelated, while the Monte Carlo simulations are for users that have a correlated CNR from time-slot to time-slot. Jakes' correlation model is used with $f_c = 1$ GHz and a user speed of $v = 30$ m/s.

As mentioned in Section 2.5, the tightness of the approximation is both influenced by the number of time-slots K and the length of a time-slot T_{TS}, for fixed f_D. The TGVP should be calculated for a relatively large number of time-slots K to obtain a tight approximation. For a fixed time-window T_W, larger K would mean shorter time-slots and thus a higher correlation between the time-slots. However, the TGVP approximation for LTE network ($K = 80$ time-slots) is better than Mobile WiMAX ($K = 16$ time-slots), showing that the number of time-slots K within the time-window T_W have a greater effect on the tightness of the approximate TGVP expression than

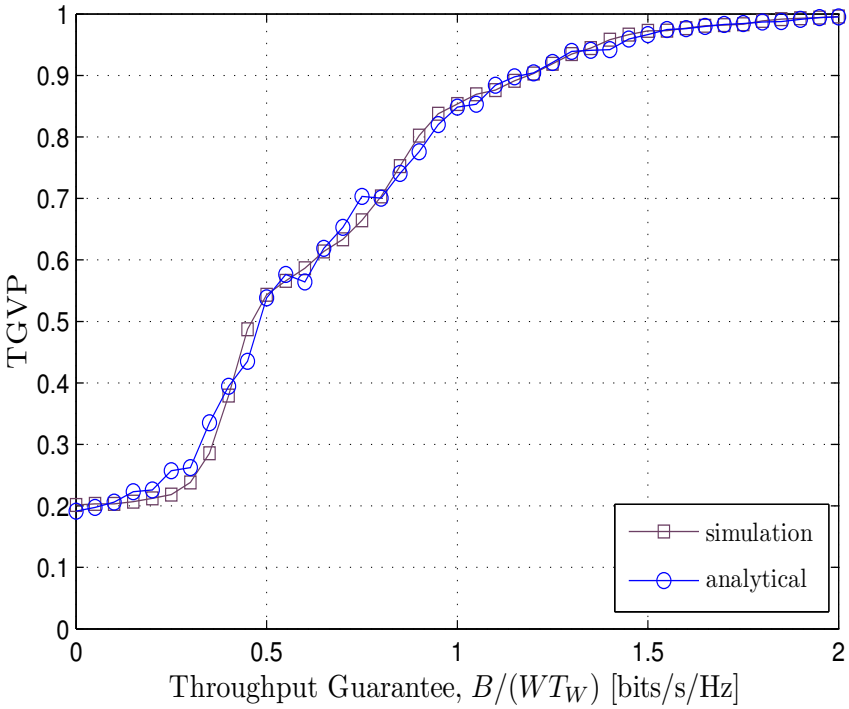


FIGURE C.1: Comparison of the approximate TGVP expression with the Monte Carlo simulated TGVP for a network with 10 users. The curves are plotted for the Mobile WiMAX network with time-slot length T_{TS} = 5 ms. A time-window of length T_W = 80 ms corresponds to $K = 16$ time-slots.

the correlation between the shorter time-slots.

These results from these curves are consistent with the results in Chapter 2. Therefore we can conclude that the derived TGVP parameters for the MTGS algorithm are correct.

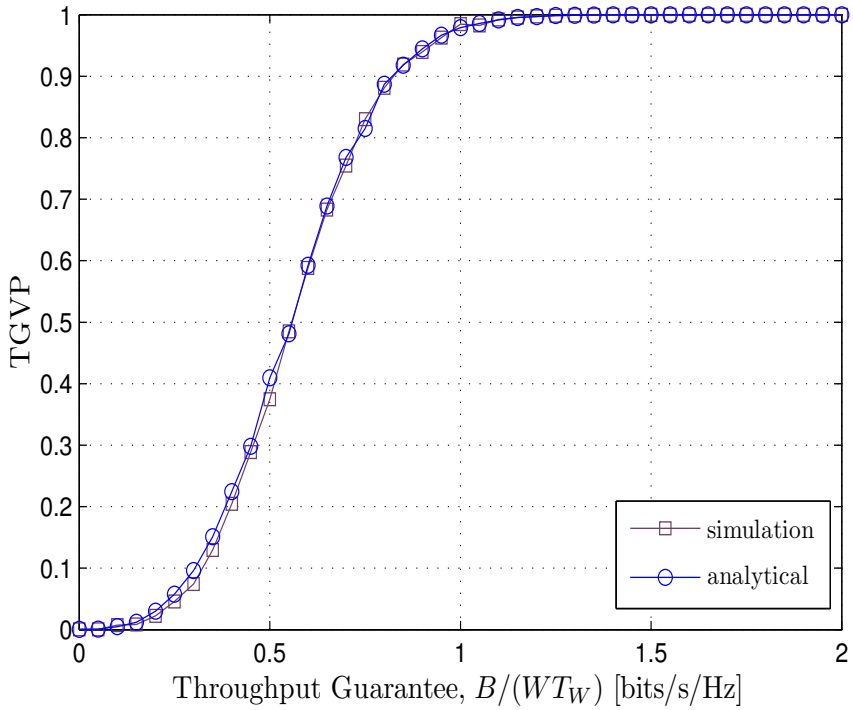


FIGURE C.2: Comparison of the approximate TGVP expression with the Monte Carlo simulated TGVP for a network with 10 users. The curves are plotted for the LTE network with time-slot length $T_{\text{TS}} = 1$ ms. A time-window of length $T_W = 80$ ms corresponds to $K = 80$ time-slots.

Appendix D

ASSE of the RR, MCS and NCS Algorithms

D.1 RR Algorithm

The ASSE of the RR scheduling algorithm simply equals the spectral efficiency averaged over all the users [Alouini and Goldsmith, 1999; Hassel, 2007, Eq. (34)]:

$$\mathrm{ASSE}_{RR} = \frac{1}{N \ln 2} \sum_{i=1}^{N} e^{1/\bar{\gamma}_i} E_1\left(\frac{1}{\bar{\gamma}_i}\right), \tag{D.1}$$

where $E_1(x)$ is the exponential integral function of first order, given in (2.19).

D.2 MCS Algorithm

The ASSE of the MCS algorithm can be expressed as [Yang and Alouini, 2006, Eq. (22)][Hassel, 2007]:

$$\mathrm{ASSE}_{MCS} = \frac{1}{\ln 2} \sum_{i=1}^{N} \frac{1}{\bar{\gamma}_i} \sum_{\tau \in T_i^N} \mathrm{sign}(\tau) \frac{e^{\left(\frac{1}{\bar{\gamma}_i}+\tau\right)}}{\frac{1}{\bar{\gamma}_i}+\tau} E_1\left(\frac{1}{\bar{\gamma}_i}+\tau\right), \tag{D.2}$$

where T_i^N denotes a set containing the terms arising from the expansion of the product $\prod_{\substack{j=1 \\ j\neq i}}^{N}(1-e^{-\gamma/\bar{\gamma}_j})$, as shown in Appendix A.

D.3 NCS Algorithm

The ASSE of the NCS algorithm can be expressed as [Yang and Alouini, 2006, Eq. (44)][Hassel, 2007]:

$$\mathrm{ASSE}_{NCS} = \frac{1}{\ln 2} \sum_{i=1}^{N} \sum_{j=0}^{N-1} \binom{N-1}{j} \frac{(-1)^j}{1+j} e^{\frac{1+j}{\bar{\gamma}_i}} E_1\left(\frac{1+j}{\bar{\gamma}_i}\right). \tag{D.3}$$

Appendix E

Fairness of the RR, MCS and NCS Algorithms

E.1 Time-Slot Fairness

E.1.1 RR Algorithm

The average number of time-slots assigned to any user is [1]

$$\mathrm{E}_K[X_{\mathrm{TS}}] = \frac{K}{N}. \tag{E.1}$$

The second moment of X_{TS} can be expressed as

$$\mathrm{E}_K[X_{\mathrm{TS}}^2] = \left(\left\lfloor \frac{K}{N} \right\rfloor\right)^2 \left(\frac{(\lfloor \frac{K}{N} \rfloor + 1)N - K}{N}\right) + \left(\left\lceil \frac{K}{N} \right\rceil\right)^2 \left(\frac{K - (\lceil \frac{K}{N} \rceil - 1)N}{N}\right) \tag{E.2}$$

for $K \bmod N \neq 0$, and

$$\mathrm{E}_K[X_{\mathrm{TS}}^2] = \left(\frac{K}{N}\right)^2, \tag{E.3}$$

otherwise. Inserting the above expressions into (3.27), a closed-form expression for the time-slot fairness of the RR algorithm can be obtained.

[1] All the expressions in this appendix have been taken from [Hassel, 2007].

E.1.2 MCS Algorithm

The expected number of time-slots assigned to an arbitrary user is

$$\mathrm{E}_K[X_{\mathrm{TS}}] = \frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K} k\, p_K(k|i) = \frac{K}{N}. \tag{E.4}$$

The corresponding second moment of the number of time-slots allocated to a user can be expressed as:

$$\mathrm{E}_K[X_{\mathrm{TS}}^2] = \frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K} k^2 p_K(k|i). \tag{E.5}$$

Inserting the expressions for these two first moments into (3.27), we obtain a closed-form expression for the time-slot fairness of the MCS algorithm.

E.1.3 NCS Algorithm

We can find the expected number of time-slots allocated to any user as:

$$\mathrm{E}_K[X_{\mathrm{TS}}] = \sum_{k=1}^{K} k\, p_K(k) = \frac{K}{N}. \tag{E.6}$$

Similarly, we can find the second moment of the time-slot allocation to be:

$$\mathrm{E}_K[X_{\mathrm{TS}}^2] = \sum_{k=1}^{K} k^2 p_K(k) = \frac{K(N+K-1)}{N^2}. \tag{E.7}$$

Inserting the expressions for $\mathrm{E}_K[X_{\mathrm{TS}}]$ and $\mathrm{E}_K[X_{\mathrm{TS}}^2]$ into (3.27), we obtain a closed-form expression for the time-slot fairness for the NCS algorithm.

E.2 Throughput Fairness

E.2.1 RR Algorithm

The first moment of the throughput allocation for the RR algorithm can be written as follows:

$$\begin{aligned}\mathrm{E}_K[X_{\mathrm{TP}}] &= \frac{1}{N}\sum_{i=1}^{N}\sum_{k=0}^{K} p_K(k)\int_0^{\infty} k\log_2(1+\gamma) f_{\gamma^*}(\gamma|i)d\gamma \\ &= \frac{K}{N^2\ln 2}\sum_{i=1}^{N} e^{1/\bar{\gamma}_i} E_1\left(\frac{1}{\bar{\gamma}_i}\right),\end{aligned} \tag{E.8}$$

where $E_1(x)$ is the exponential integral function of first order given in (2.19). Furthermore, the second moment of the throughput allocation can be expressed as:

$$\begin{aligned} \mathrm{E}_K[X_{\mathrm{TP}}^2] &= \frac{1}{N}\sum_{i=1}^{N}\sum_{k=0}^{K} p_K(k)\int_0^{\infty}(k\log_2(1+\gamma))^2 f_{\gamma^*}(\gamma|i)d\gamma \\ &= \frac{\mathrm{E}_K[X_{\mathrm{TS}}^2]}{N(\ln 2)^2}\sum_{i=1}^{N}\frac{1}{\bar{\gamma}_i}\Psi\left(\frac{1}{\bar{\gamma}_i}\right), \end{aligned} \tag{E.9}$$

where $\mathrm{E}_K[X_{\mathrm{TS}}^2]$ is the second moment of the time-slot allocation for RR and $\Psi(x)$ is given in (2.21). Inserting the obtained expressions for these moments of the throughput allocation into (3.30), we obtain a closed-form expression for the throughput fairness of the RR algorithm.

E.2.2 MCS Algorithm

The first moment of the throughput allocation for the MCS algorithm can now be found as follows:

$$\begin{aligned} \mathrm{E}_K[X_{\mathrm{TP}}] &= \frac{1}{N}\sum_{i=1}^{N}\sum_{k=0}^{K} p_K(k|i)\int_0^{\infty} k\log_2(1+\gamma) f_{\gamma^*}(\gamma|i)d\gamma \\ &= \frac{1}{N\ln 2}\sum_{i=1}^{N}\sum_{k=0}^{K} k\frac{p_K(k|i)}{\bar{\gamma}_i p(i)}\sum_{\tau\in T_i^N}\mathrm{sign}(\tau)\frac{e^{\left(\frac{1}{\bar{\gamma}_i}+\tau\right)}}{\frac{1}{\bar{\gamma}_i}+\tau}E_1\left(\frac{1}{\bar{\gamma}_i}+\tau\right). \end{aligned} \tag{E.10}$$

Similarly, we can obtain the second moment of the throughput allocation as:

$$\begin{aligned} \mathrm{E}_K[X_{\mathrm{TP}}^2] &= \frac{1}{N}\sum_{i=1}^{N}\sum_{k=0}^{K} p_K(k|i)\int_0^{\infty}[k\log_2(1+\gamma)]^2 f_{\gamma^*}(\gamma|i)d\gamma \\ &= \frac{1}{N(\ln 2)^2}\sum_{i=1}^{N}\sum_{k=0}^{K} k^2\frac{p_K(k|i)}{\bar{\gamma}_i p(i)}\sum_{\tau\in T_i^N}\mathrm{sign}(\tau)\Psi\left(\frac{1}{\bar{\gamma}_i}+\tau\right). \end{aligned} \tag{E.11}$$

Inserting the expressions for $\mathrm{E}_K[X_{\mathrm{TP}}]$ and $\mathrm{E}_K[X_{\mathrm{TP}}^2]$ into (3.30), we obtain a closed-form expression for the throughput fairness of the MCS algorithm.

E.2.3 NCS Algorithm

We can find the first moment of the throughput allocation as:

$$\begin{aligned} \mathrm{E}_K[X_{\mathrm{TP}}] &= \frac{1}{N}\sum_{i=1}^{N}\sum_{k=0}^{K} p_K(k)\int_0^{\infty} k\log_2(1+\gamma) f_{\gamma^*}(\gamma|i)d\gamma \\ &= \frac{K}{N\ln 2}\sum_{i=1}^{N}\sum_{j=0}^{N-1}\binom{N-1}{j}\frac{(-1)^j}{1+j}e^{\frac{1+j}{\bar{\gamma}_i}}E_1\left(\frac{1+j}{\bar{\gamma}_i}\right). \end{aligned} \quad \text{(E.12)}$$

Similarly, we can obtain the second moment of the throughput allocation as:

$$\begin{aligned} \mathrm{E}_K[X_{\mathrm{TP}}^2] &= \frac{1}{N}\sum_{i=1}^{N}\sum_{k=0}^{K} p_K(k)\int_0^{\infty} [k\log_2(1+\gamma)]^2 f_{\gamma^*}(\gamma|i)d\gamma \\ &= \frac{K(N+K-1)}{(N\ln 2)^2}\sum_{i=1}^{N}\frac{1}{\bar{\gamma}_i}\sum_{j=0}^{N-1}\binom{N-1}{j}(-1)^j\Psi\left(\frac{1+j}{\bar{\gamma}_i}\right). \end{aligned} \quad \text{(E.13)}$$

As for the RR and MCS algorithms, the closed-form throughput fairness expression for NCS can subsequently be found by inserting these two moments into (3.30).

Appendix F

Derivation of (4.14)

As stated in Section 4.2, due to imperfect channel prediction, there is always a possibility that the predicted CNR $\hat{\gamma}_i$ of user i turns out to be greater than the actual CNR γ_i, and sending the data at the rate $\hat{r}_i = W \log_2(1 + \hat{\gamma}_i)$ would result in outage. We therefore introduced a back-off CNR γ_b ($< \hat{\gamma}_i$) and send the data at a rate $W \log_2(1 + \gamma_b)$. Still, the outage will not be avoided completely. It is, therefore, desirable to control the probability of this event, i.e. to demand

$$\Pr[\gamma_i < \gamma_b | \hat{\gamma}_i] \leq \epsilon, \tag{F.1}$$

where ϵ is a small constant chosen by the system designer. The probability of the complementary event is given as [Jetlund et al., 2004; Øien et al., 2002]

$$\Pr[\gamma_i > \gamma_b | \hat{\gamma}_i] = 1 - \epsilon. \tag{F.2}$$

If we assume Rayleigh faded channel gains and MAP-optimal prediction, the actual and predicted CNRs are correlated with normalized correlation ρ_i, and have average CNRs $\bar{\gamma}_i$ and $\rho_i \bar{\gamma}_i$, respectively. The two then follow a joint exponential PDF given as follows [Nagao and Kadoya, 1971]:

$$f_{\gamma,\hat{\gamma}}(\gamma, \hat{\gamma}) = \frac{1}{\rho \bar{\gamma}^2 (1-\rho)} I_0 \left(\frac{2\sqrt{\gamma \hat{\gamma}}}{\bar{\gamma}(1-\rho)} \right) e^{-\left(\frac{\gamma}{\bar{\gamma}(1-\rho)} + \frac{\hat{\gamma}}{\rho \bar{\gamma}(1-\rho)} \right)},$$

where $I_0(\cdot)$ is the zeroth-order modified Bessel function of first kind. We have dropped the subscript i for simplification. The conditional PDF of the

actual and predicted CNR can be obtained using Bayes' rule as follows:

$$\begin{aligned} f_{\gamma|\hat{\gamma}}(\gamma|\hat{\gamma}) &= \frac{f_{\gamma,\hat{\gamma}}(\gamma,\hat{\gamma})}{f_{\hat{\gamma}}(\hat{\gamma})} \\ &= \frac{1}{\bar{\gamma}(1-\rho)} I_0\left(\frac{2\sqrt{\gamma\hat{\gamma}}}{\bar{\gamma}(1-\rho)}\right) e^{-\left(\frac{\gamma}{\bar{\gamma}(1-\rho)}+\frac{\hat{\gamma}}{\bar{\gamma}(1-\rho)}\right)}. \end{aligned}$$

The probability in (F.2) can now be obtained as follows:

$$\begin{aligned} \Pr[\gamma_i > \gamma_b|\hat{\gamma}_i] &= \int_{\gamma_b}^{\infty} f_{\gamma|\hat{\gamma}}(\gamma|\hat{\gamma})d\gamma \\ &= \int_{\gamma_b}^{\infty} \frac{1}{\bar{\gamma}(1-\rho)} I_0\left(\frac{2\sqrt{\gamma\hat{\gamma}}}{\bar{\gamma}(1-\rho)}\right) e^{-\left(\frac{\gamma}{\bar{\gamma}(1-\rho)}+\frac{\hat{\gamma}}{\bar{\gamma}(1-\rho)}\right)} d\gamma \\ &= \int_{\gamma_b}^{\infty} \frac{1}{\bar{\gamma}(1-\rho)} I_0\left(\sqrt{\frac{2\gamma}{\bar{\gamma}(1-\rho)}}\sqrt{\frac{2\hat{\gamma}}{\bar{\gamma}(1-\rho)}}\right) \\ &\quad \times e^{-\frac{1}{2}\left(\frac{2\gamma}{\bar{\gamma}(1-\rho)}+\frac{2\hat{\gamma}}{\bar{\gamma}(1-\rho)}\right)} d\gamma. \end{aligned} \tag{F.3}$$

Then we make the following substitutions and change of the integration variables in (F.3):

$$a = \sqrt{\frac{2\hat{\gamma}}{\bar{\gamma}(1-\rho)}}, b = \sqrt{\frac{2\gamma_b}{\bar{\gamma}(1-\rho)}}, \text{ and } x = \sqrt{\frac{2\gamma}{\bar{\gamma}(1-\rho)}}.$$

This will result in

$$\begin{aligned} \Pr(\gamma_i > \gamma_b|\hat{\gamma}_i) &= \int_b^{\infty} x I_0(ax) e^{-\frac{1}{2}(a^2+x^2)} dx \\ &= Q(a,b), \end{aligned} \tag{F.4}$$

where $Q(a,b)$ is the Marcum-Q function [Marcum, 1948].

Bibliography

Alouini, M.-S. and A. J. Goldsmith (1999, July).
Capacity of Rayleigh fading channels under different adaptive transmission and diversity-combining techniques.
IEEE Transactions on Vehicular Technology 48(4), 1165–1181.

Alouini, M.-S. and A. J. Goldsmith (2000, May).
Adaptive modulation over Nakagami fading channels.
Kluwer Journal on Wireless Communications 13, 119–143.

Alouini, M. S. and M. K. Simon (2000, March).
An MGF-based performance analysis of generalized selection combining over Rayleigh fading channels.
IEEE Transactions on Communications 48(3), 401–415.

Andrews, M., K. Kumaran, K. Ramanan, A. Stolyar, R. Vijayakumar, and P. Whiting (2000, April).
CDMA data QoS scheduling on the forward link with variable channel conditions.
Bell Labs Technical Memorandum.

Andrews, M., K. Kumaran, K. Ramanan, A. Stolyar, P. Whiting, and R. Vijayakumar (2001, February).
Providing quality of service over a shared wireless link.
IEEE Communications Magazine, 150–154.

Andrews, M., L. Qian, and A. Stolyar (2005, March).
Optimal utility based multi-user throughput allocation subject to throughput constraints.

In *Proceedings of the 24th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM'05)*, Volume 4, Miami, FL, USA, pp. 2415–2424.

Avidor, D., S. Mukherjee, J. Ling, and C. Papadias (2004, September).
On some properties of the proportional fair scheduling policy.
In *Proceedings of IEEE Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC'04)*, Volume 2, Barcelona, Spain, pp. 853–858.

Bang, H. J., T. Ekman, and D. Gesbert (2008, February).
Channel predictive proportional fair scheduling.
IEEE Transactions on Wireless Communications 7(2), 482–487.

Borst, S. C. and P. Whiting (2003, May).
Dynamic channel-sensitive scheduling algorithms for wireless data throughput optimization.
IEEE Transactions on Vehicular Technology 52(3), 569–586.

Bucklew, J. A. (2004).
Introduction to Rare Event Simulation.
Springer Series in Statistics. New York: Springer Verlag.

Caire, G. and S. Shamai (2003, July).
On the achievable throughput of a multiantenna Gaussian broadcast channel.
IEEE Transactions on Information Theory 49(7), 1691–1706.

Cardei, M., M. T. Thai, Y. Li, and W. Wu (2005, March).
Energy-efficient target coverage in wireless sensor networks.
In *Proceedings of IEEE Joint Conference of the Computer and Communications Societies (INFOCOM'05)*, Miami, USA, pp. 1976–1984.

Cavers, J. K. (1972, February).
Variable-rate transmission for Rayleigh fading channels.
IEEE Transactions on Communications COM-20(1), 74–83.

Chaponniere, E. F., P. J. Black, J. M. Holtzman, and D. N. C. Tse (2002, September).
Transmitter directed code division multiple access system using path diversity to equitably maximize throughput.

U.S. Patent #6449490, http://www.eecs.berkeley.edu/~dtse/scheduler_patent.html.

Chen, N. and S. Jordan (2009, February).
Downlink scheduling with guarantees on the probability of short-term throughput.
IEEE Transactions on Wireless Communications 8(2), 593–598.

Cohen, H. (1993).
A Course in Computational Algebraic Number Theory.
New York: Springer Verlag.

Conte, E., S. Tomasin, and N. Benvenuto (2008, September).
Scheduling strategies for multiuser MIMO OFDM systems with limited feedback.
In *Proceedings of IEEE Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC'08)*, Cannes, pp. 1–5.

Duong, D. V. and G. E. Øien (2007, March).
Optimal pilot spacing and power in rate-adaptive MIMO diversity system with imperfect CSI.
IEEE Transactions on Wireless Communications 6(3), 845–851.

Eng, T., N. Kong, and L. B. Milstein (1996, September).
Comparison of diversity combining techniques for Rayleigh-fading channels.
IEEE Transactions on Communications 44(9), 1117–1129.

Eric W. Weisstein.
Cumulant.
http://mathworld.wolfram.com/Cumulant.html.
From MathWorld–A Wolfram Web Resource.

Eric W. Weisstein.
Cumulant-Generating Function.
http://mathworld.wolfram.com/Cumulant-GeneratingFunction.html.
From MathWorld–A Wolfram Web Resource.

Eric W. Weisstein.
Generalized Hypergeometric Function.

http://mathworld.wolfram.com/GeneralizedHypergeometricFunction.html.
From MathWorld–A Wolfram Web Resource.

Eric W. Weisstein.
Incomplete Gamma Function.
http://mathworld.wolfram.com/IncompleteGammaFunction.html.
From MathWorld–A Wolfram Web Resource.

Eric W. Weisstein.
Regularized Hypergeometric Function.
http://mathworld.wolfram.com/RegularizedHypergeometricFunction.html.
From MathWorld–A Wolfram Web Resource.

Fang, Y. and Y. Zhang (2002, March).
Call admission control schemes and performance analysis in wireless mobile networks.
IEEE Transactions on Vehicular Technology 51(2), 371–382.

Fattah, H. and C. Leung (2002, October).
An overview of scheduling algorithms in wireless multimedia networks.
IEEE Wireless Communications 9(5), 76–83.

Ferrari, D. (1990, November).
Client requirements for real-time communication services.
IEEE Transactions on Communications 28, 65–72.

Foschini, G. J. and M. J. Gans (1998, March).
On limits of wireless communication in a fading environment when using multiple antennas.
Wireless Personal Communications Journal 6, 311–335.

Freescale Semiconductor (2007, July).
Overview of the 3GPP Long Term Evolution Physical Layer.
http://www.freescale.com/files/wireless_comm/doc/white_paper/3GPPEVOLUTIONWP.pdf.

Ghaderi, M. and R. Boutaba (2006, February).
Call admission control in mobile cellular networks: a comprehensive survey.
Wireless Communications & Mobile Computing 6(1), 69–93.

Girici, T., C. Zhu, J. R. Agre, and A. Ephremides (2010, February).
Proportional fair scheduling algorithm in OFDMA-based wireless systems with QoS constraints.
Journal of Communications and Networks 12(1), 30–42.

Gjendemsjø, A., G. E. Øien, H. Holm, M.-S. Alouini, D. Gesbert, K. J. Hole, and P. Orten (2008).
Rate and power allocation for discrete-rate link adaptation.
EURASIP Journal on Wireless Communications and Networking 2008.
Article ID 394124, 11 pages.

Goldsmith, A. (2005).
Wireless Communications.
Cambridge University Press.

Goldsmith, A. J. and S.-G. Chua (1997, October).
Variable-rate variable-power M-QAM for fading channels.
IEEE Transactions on Communications COM-45, 1218–1230.

Goldsmith, A. J. and P. P. Varaiya (1997, November).
Capacity of fading channels with channel side information.
IEEE Transactions on Information Theory IT-43(6), 1896–1992.

Gong, X., S. Vorobyov, and C. Tellambura (2011, April).
Joint bandwidth and power allocation with admission control in wireless multi-user networks with and without relaying.
IEEE Transactions on Signal Processing 59, 1801–1813.

Gradshteyn, I. S. and I. M. Ryzhik (2007).
Table of Integrals, Series, and Products (7th ed.).
San Diego, CA, USA: Academic Press.

Hammarwall, D., M. Bengtsson, and B. Ottersten (2008, March).
Acquiring partial CSI for spatially selective transmission by instantaneous channel norm feedback.
IEEE Transactions on Signal Processing 56, 1188–1204.

Hassel, V. (2007).
Design Issues and Performance Analysis for Opportunistic Scheduling Algorithms in Wireless Networks.

Ph. D. thesis, Norwegian University of Science and Technology, Trondheim.
http://ntnu.diva-portal.org/smash/record.jsf?pid=diva2:121874.

Hassel, V., S. de la Kethulle de Ryhove, and G. E. Øien (2007, October).
Scheduling algorithms for increased throughput guarantees in wireless networks.
In *Proceedings of IEEE International Symposium on Wireless Communication Systems (ISWCS'07)*, Trondheim, Norway, pp. 401–406.

Hassel, V., M. R. Hanssen, and G. E. Øien (2006, June).
Spectral efficiency and fairness for opportunistic round robin scheduling.
In *Proceedings of IEEE International Conference on Communications (ICC'06)*, Volume 2, Istanbul, Turkey, pp. 784–789.

Hassel, V., G. E. Øien, and D. Gesbert (2007, December).
Throughput guarantees for wireless networks with opportunistic scheduling: A comparative study.
IEEE Transactions on Wireless Communications 6(12), 4215–4220.

Hasu, V. (2007).
Radio resource management in wireless communication: beamforming, transmission power control, and rate allocation.
Ph. D. thesis, Helsinki University of Technology, Finland.
http://lib.tkk.fi/Diss/2007/isbn9789512288229.

Hole, K. J., H. Holm, and G. E. Øien (2000, July).
Adaptive multi-dimensional coded modulation over flat fading channels.
IEEE Journal on Selected Areas in Communications 18(7), 1153–1158.

Holm, H. (2002).
Adaptive coded modulation performance and channel estimation tools for flat fading channels.
Ph. D. thesis, Norwegian University of Science and Technology, Trondheim.
http://www.henrikholm.com/avhandling/thesis.pdf.

Holm, H., G. Øien, M.-S. Alouini, D. Gesbert, and K. J. Hole (2003, June).

Optimal design of adaptive coded modulation schemes for maximum average spectral efficiency.
In *Proceedings of 4th IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC'03)*, Rome, Italy, pp. 403–407.

Holtzman, J. M. (2001, September).
Asymptotic analysis of proportional fair algorithm.
In *Proceedings of IEEE Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC'01)*, Volume 2, San Diego, CA, USA, pp. F–33–F–37.

Hugo Touchette (2007, August).
Legendre-Fenchel transforms in a nutshell.
http://www.maths.qmul.ac.uk/~ht/archive/lfth2.pdf.

Jagannathan, K. P., S. Borst, P. Whiting, and E. Modiano (2006, April).
Efficient scheduling of multi-user multi-antenna systems.
In *International Symposium on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (WiOpt'06)*, Boston, MA, USA, pp. 1–8.

Jain, R., D. Chiu, and W. Hawe (1984, September).
A quantitative measure of fairness and discrimination for resource allocation in shared computer systems.
DEC Research Report TR-301, Digital Equipment Corporation, Maynard, MA, USA.

Jetlund, O., G. E. Øien, H. Holm, and K. J. Hole (2004, August).
Spectral efficiency bounds for adaptive coded modulation with outage probability constraints and imperfect channel prediction.
In *Proceedings of Nordic Radio Symposium*, Oulu, Finland.

Ji, Z., Y. Yang, J. Zhou, M.Takai, and R. Bagrodia (2004, September).
Exploiting medium access diversity in rate adaptive wireless LANs.
In *Proceedings of ACM International Conference on Mobile Computing and Networking (MOBICOM'04)*, Philadephia, PA, USA, pp. 345–359.

Jindal, N. and A. Goldsmith (2005, May).
Dirty-paper coding versus TDMA for MIMO broadcast channels.
IEEE Transactions on Information Theory 51(5), 1783–1794.

Johansson, M. (2004, July).

Diversity-enhanced equal access - considerable throughput gains with 1-bit feedback.
In *Proceedings of IEEE Workshop on Signal Processing Advances in Wireless Communications (SPAWC'04)*, Lisbon, Portugal, pp. 6–10.

Julian, D., M. Chiang, D. O'Neill, and S. Boyd (2002, June).
Qos and fairness constrained convex optimization of resource allocation for wireless cellular and ad hoc networks.
In *Proceedings of IEEE Joint Conference of the Computer and Communications Societies (INFOCOM'02)*, New York, USA, pp. 477—-486.

K. Aho, K., I. Repo, J. Puttonen, T. Henttonen, M. Moisio, J. Kurjenniemi, and K. Chang (2010, May).
Benchmarking of VoIP over HSDPA and LTE performance with realistic network data.
In *Proceedings of the 5th IEEE International Symposium on Wireless Pervasive Computing (ISWPC'10)*, Modena, Italy, pp. 401–406.

Kastrinogiannis, T. and S. Papavassiliou (2007, April).
Probabilistic short-term delay and throughput requirements of multimedia services in high throughput wireless networks.
In *Proceedings of IEEE Sarnoff Symposium on Advances in Wired and Wireless Communications*, Nassau Inn, Princeton, NJ, pp. 1–5.

Kastrinogiannis, T. and S. Papavassiliou (2008, November).
Utility based short-term throughput driven scheduling approach for efficient resource allocation in CDMA wireless networks.
Wireless Personal Communications Journal, Springer 52(3), 517–535.

Kiusalaas, J. (2010).
Numerical Methods in Engineering with MATLAB (2nd ed.).
Cambridge University Press.

Knopp, R. and P. A. Humblet (1995, June).
Information capacity and power control in single cell multiuser communications.
In *Proceedings of IEEE International Conference on Communications (ICC'95)*, Seattle, WA, USA, pp. 331–335.

Kountouris, M. and D. Gesbert (2005, September).

Memory-based opportunistic multi-user beamforming.
In *Proceedings of International Symposium on Information Theory (ISIT'05)*, Adelaide, SA, pp. 1426–1430.

Kreyszig, E. (1993).
Advanced Engineering Mathematics (7th ed.).
New York: John Wiley & Sons, Inc.

Kulkarni, S. S. and C. Rosenberg (2003, December).
Opportunistic scheduling policies for wireless systems with short term fairness constraints.
In *Proceedings of IEEE Global Communications Conference (GLOBECOM'03)*, San Francisco, CA, USA, pp. 533–537.

Kumaran, K. and H. Viswanathan (2005, May).
Joint power and bandwidth allocation in downlink transmission.
IEEE Transactions on Wireless Communications 4, 1008–1016.

Lawrey, E. (1999, August).
Multiuser OFDM.
In *Proceedings of 5th International Symposium on Signal Processing and its Applications (ISSPA'99)*, Brisbane, Australia, pp. 761–764.

Liu, X., E. K. P. Chong, and N. B. Shroff (2003).
A framework for opportunistic scheduling in wireless networks.
In *Computer Networks*, Volume 41, pp. 451–474.

Madkour, M. F., S. C. Gupta, and Y.-P. E. Wang (2002, January).
Successive interference cancellation algorithms for downlink W-CDMA communications.
IEEE Transactions on Wireless Communications 1(1), 169–177.

Marcum, J. I. (1948).
A statistical theory of target detection by pulsed radar: Mathematical appendix.
In *RAND Corporation, California, Research Memo RM-753*, California, USA.

Maruddani, B. and A. Kurniawan (2010, July).
Power control and diversity performance analysis in CDMA systems.
In *Proceedings of Electromagnetics Research Symposium*, Cambridge, USA, pp. 101–106.

Miao, G.-W., N. Himayat, G. Y. Li, and A. Swami (2009, April).
Cross-layer optimization for energy-efficient wireless communications: A survey.
Wiley Journal of Wireless Communications and Mobile Computing 9(4), 529–542.

Mostofi, Y. and D. C. Cox (2006, February).
Mathematical analysis of the impact of timing synchronization errors on the performance of an OFDM system.
IEEE Transactions on Communications 54(2), 226–230.

Nagao, M. and M. Kadoya (1971, March).
Two-variate exponential distributions and its numerical table for engineering applications.
Bulletin of the Disaster Prevention Research Institute 20(178), 183–215.

Nguyen, T. D. and Y. Han (2006, November).
A proportional fairness algorithm with QoS provision in downlink OFDMA systems.
IEEE Communications Letters 10(11), 760–762.

Nomor Research: White Paper (2006, October).
Technology of High Speed Packet Access (HSPA).
http://www.nomor.de/uploads/b0/2m/b02mwrVvIa5ZUtVrFeSP1w/Technology_of_HSPA.pdf.

Øien, G. E., H. Holm, and K. J. Hole (2002, September).
Adaptive coded modulation with imperfect channel state information: System design and performance analysis aspects.
In *Proceedings of IEEE International Symposium on Advances in Wireless Communications (ISWC'02)*, Victoria, BC, Canada.

Øien, G. E., H. Holm, and K. J. Hole (2004, May).
Impact of channel prediction on adaptive coded modulation performance in Rayleigh fading.
IEEE Transactions on Vehicular Technology 53(3), 758–769.

Phan, K. T., T. Le-Ngoc, S. A. Vorobyov, and C. Tellambura (2009, May).
Power allocation in wireless multi-user relay networks.
IEEE Transactions on Wireless Communications 8, 2535–2545.

Rappaport, T. S. (2002).
Wireless Communications: Principles and Practice (2nd ed.).
New Jersey: Prentice-Hall, Inc.

Rasool, J., V. Hassel, S. de la Kethulle de Ryhove, and G. E. Øien (2011).
Opportunistic scheduling policies for improved throughput guarantees in wireless networks.
EURASIP Journal on Wireless Communications and Networking.
doi:10.1186/1687-1499-2011-43, 12 pages.

Rasool, J. and G. E. Øien (2009, November).
Multiuser MIMO systems for spectrally efficient communications in future aeronautical services.
In *Proceedings of VERDIKT Conference 2009*, Oslo, Norway.

Rasool, J. and G. E. Øien (2010a, November).
Improving the throughput guarantees offered in wireless networks.
In *Proceedings of VERDIKT Conference 2010*, Oslo, Norway.

Rasool, J. and G. E. Øien (2010b, April).
A simplified adaptive scheduling algorithm for increased throughput guarantees.
In *Proceedings of 2010 European Wireless Conference (EW2010)*, Lucca, Italy, pp. 182–189.

Rasool, J. and G. E. Øien (2011a, March).
Quantifying the throughput guarantees offered in wireless networks.
In *Proceedings of IEEE Wireless Communications and Networking Conference (WCNC'11)*, Cancun, Mexico, pp. 1173–1178.

Rasool, J. and G. E. Øien (2011b, March).
Scheduling policies for improved throughput guarantees in MIMO broadcast channel.
In *Proceedings of NEWCOM++/COST2100 Joint Workshop on Wireless Communications (JNCW'11)*, Paris, France, pp. 391–395.

Rasool, J. and G. E. Øien (2011c, June).
Scheduling schemes for improved throughput guarantees in MIMO broadcast channels.

In *Proceedings of 12th IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC'11)*, San Francisco, USA, pp. 391–395.

Rasool, J. and G. E. Øien (2012a, September).
Joint power and bandwidth allocation for discrete-rate multi-user link adaptation with imperfect channel state information.
Submitted to *IEEE Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC'12)*, Sydney, Australia.

Rasool, J. and G. E. Øien (2012b, April).
Maximizing the throughput guarantees in wireless networks under imperfect channel knowledge.
Accepted for *IEEE Wireless and Networking Conference (WCNC'12)*, Paris, France.

Rasool, J., G. E. Øien, J. E. Håkegård, and T. A. Myrvoll (2009, September).
On multiuser MIMO capacity benefits in air-to-ground communication for airport traffic management.
In *Proceedings of International Symposium on Wireless Communication Systems (ISWCS' 09)*, Siena, Italy, pp. 458–462.

Redana, S., A. Frediani, and A. Capone (2008, September).
Quality of service scheduling based on utility prediction.
In *Proceedings of IEEE Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC'08)*, Cannes, pp. 1–5.

Ridders, C. (1979, November).
A new algorithm for computing a single root of a real continuous function.
IEEE Transactions on Circuits and Systems 26, 979–980.

S. Lu, V. B. and R. Srikant (1999, August).
Fair scheduling in wireless packet networks.
IEEE/ACM Transactions on Networking 7(4), 473–489.

Sharif, M. and B. Hassibi (2005, February).
On the capacity of MIMO broadcast channels with partial side information.
IEEE Transactions on Information Theory 51(2), 506–522.

Simon, M. K. and M.-S. Alouini (2005).
Digital communication over fading channels (2nd ed.).
New York: John Wiley & Sons, Inc.

Stemick, M. and H. Rohling (2007, April).
Effect of carrier frequency offset on the channel capacity in multi user OFDM-FDMA systems.
In *Proceedings of 4th Workshop of COST 289*, Göteborg, Sweden.

Sternad, M., T. Svenson, and G. Klang (2006, June).
WINNER MAC for cellular transmission.
In *Proceedings of IST Mobile Summit*, Mykonos, Greece.

Stolyar, A. L. and K. Ramanan (2001).
Largest weighted delay first scheduling: Large deviations and optimality.
The Annals of Applied Probability 11(1), 1–48.

Sun, J., D. Howie, and J. Sauvola (2001, November).
Mobility management techniques for the next generation wireless networks.
In *Proceedings of SPIE Asia-Pacific Optical and Wireless Communications Conference*, Beijing, China, pp. 155–166.

Sung, C. W. and W. S. Wong (2001, July).
Power control and rate management for wireless multimedia CDMA systems.
IEEE Transactions on Communications 49(7), 1215–1226.

Svensson, A., G. E. Øien, M. S. Alouini, and S. Sampei (2007, December).
Special issue on adaptive modulation and transmission in wireless systems.
In *Proceedings of IEEE*, pp. 2269–2273.

Tao, T. and A. Czylwik (2011, February).
Performance analysis of link adaptation in LTE systems.
In *Proceedings of International ITG Workshop on Smart Antennas (WSA'11)*, Aachen, Germany, pp. 1–5.

Tse, D. and P. Viswanath (2005).

Fundamentals of Wireless Communication.
Cambridge University Press.

Vicario, J. L., R. Bosisio, C. Anton-Haro, and U. Spagnolini (2008, September).
Beam selection strategies for orthogonal random beamforming in sparse networks.
IEEE Transactions on Wireless Communications 7(9), 3385–3396.

Viswanath, P. and D. N. C. Tse (2003, August).
Sum capacity of the vector Gaussian broadcast channel and uplink-downlink duality.
IEEE Transactions on Information Theory 49(8), 1912–1921.

Viswanath, P., D. N. C. Tse, and R. Laroia (2002, June).
Opportunistic beamforming using dumb antennas.
IEEE Transactions on Information Theory 48(6), 1277–1294.

Webb, W. T. and R. Steele (1995, July).
Variable-rate QAM for mobile radio.
IEEE Transactions on Communications 43(7), 2223–2230.

WiMAX Forum (2006, May).
Mobile WiMAX – Part II: A Comparative Analysis.
http://www.wimaxforum.org/technology/downloads/Mobile_WiMAX_Part2_Comparative_Analysis.pdf.

Wolfram Research Inc.
Generalized incomplete gamma function: Primary definition.
http://functions.wolfram.com/06.07.02.0001.01.

Wolfram Research Inc.
Generalized incomplete gamma function: Representation through equivalent functions: With related functions.
http://functions.wolfram.com/06.07.27.0002.01.

Wolfram Research Inc.
Incomplete gamma function: Differentiation : Low-order differentiation.
http://functions.wolfram.com/GammaBetaErf/Gamma2/20/01/01/.

Wong, C. Y., R. S. Cheng, K. B. Letaief, and R. D. Murch (1999, October).
Multiuser OFDM with adaptive subcarrier, bit, and power allocation.
IEEE Journal on Selected Areas in Communications 17(10), 1747–1758.

Yang, L. and M.-S. Alouini (2006, November).
Performance analysis of multiuser selection diversity.
IEEE Transactions on Vehicular Technology 55(6), 1848–1861.

Zander, J. (1997, August).
Radio resource management in future wireless networks: Requirements and limitations.
IEEE Communications Magazine 35(8), 30–36.

Zhang, Y., H. Hu, and M. Fujise (2007).
Resource, Mobility and Security Management in Wireless Networks and Mobile Communications (1st ed.).
Auerbach Publications.

Zhang, Z., Y. He, and E. K. P. Chong (2005, March).
Opportunistic downlink scheduling for multiuser OFDM systems.
In *Proceedings of IEEE Wireless Communications and Networking Conference (WCNC'05)*, New Orleans, USA, pp. 1206–1212.

Zheng, Y. R. and C. Xiao (2003, June).
Simulation models with correct statistical properties for Rayleigh fading channels.
IEEE Transactions on Communications 51(6), 920–928.

ISBN 978-82-471-3396-5 (printed ver.)
ISBN 978-82-471-3397-2 (electronic ver.)
ISSN 1503-8181

NTNU – Trondheim
Norwegian University of
Science and Technology