Some Cross-Layer Design and Performance Issues in Wireless Networks

by

Md. Jahangir Hossain

M.A.Sc., The University of Victoria, Victoria, BC, 2003 B.Sc., Bangladesh University of Engineering and Technology, 2000

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF

Doctor of Philosophy

in

The Faculty of Graduate Studies

(Electrical and Computer Engineering)

The University Of British Columbia

October, 2007

© Md. Jahangir Hossain, 2007

Abstract

Present and future generation wireless systems are designed to support a wide variety of services, including video conferencing, media streaming, Internet access, email transfer, remote file access and web-browsing. Different applications have different traffic characteristics and quality of service (QoS) requirements. These QoS parameters include delay constraint, packet loss rate (PLR) and packet error rate (PER) requirement. For example, depending on the throughput or delay requirements of the services, wireless data services have already been classified as guaranteed or best-effort services. Mobile wireless links are time-varying in nature and error prone. In addition, communication resources, e.g., transmit power and bandwidth, are scarce. Through a cross-layer design approach, these inherent quality parameters of different services can facilitate efficient use of limited communication resources. In particular, the relative delay tolerance of data applications, in conjunction with the bursty activity patterns and time-varying nature of the wireless channels, opens up the possibility of scheduling transmissions to efficiently allocate limited wireless communication resources.

The main contribution of this dissertation is the design of innovative scheduling schemes through *cross-layer design approach* that jointly considers the time-varying nature of wireless channels as well as the required QoS parameters of different services. This dissertation makes four major contributions, as follows. First, we develop modulation-assisted two-user opportunistic scheduling schemes. The novelty of the proposed schemes is improved delay performance and fairness for the serviced users in wireless networks without degradation of the system spectral efficiency (SE), compared with the classical single best user scheduling (SBS) scheme. Second, we develop an adaptive scheduling protocol for multi-resolution image/video data transmission over time-varying channels. The protocol is optimized for real-time multi-resolution data. Third, we devise a novel variance-bounded rate adaptation (VBRA) scheme for delay-constrained communications. We explore an application of the proposed VBRA scheme for media streaming over time-varying channels in order to minimize the transmit power while maintaining a low play-out buffer outage probability. Fourth, we develop link adaptation algorithms for a multicarrier system for generalized delay-constrained communication over time-varying channels.

iii

Table of Contents

Abstract	ii
Table of Contents	iv
List of Tables	x
List of Figures	xi
List of Abbreviations	xv
Acknowledgements	viii
Dedication	kix
Statement of Co-Authorship	xx
1 Introduction	1
1.1 General Background and Motivation	2
1.1.1 Opportunistic scheduling	2
1.1.2 Multi-resolution image/video data transmission	4
1.1.3 Link adaptation for OFDM-based systems	5
1.1.4 Variance-bounded rate adaptation for delay-constrained communi-	
cations	7
1.2 Thesis Outline	8
Bibliography	9
2 Multi-User Opportunistic Scheduling Using Power-Controlled Hierarchical Cor	1-
stellations	12

	,	
2.1	Multi-User Opportunistic Scheduling	1
	2.1.1 Hierarchical constellations	1
	2.1.2 Hierarchical constellation-based multi-user opportunistic schedul-	
	ing	10
	2.1.3 Comparison with uniform QAM constellation-based scheme	1′
2.2	Performance Analysis and Results	1′
	2.2.1 Outage probability	1
	2.2.2 Average transmit power	1
	2.2.3 Buffer density and packet loss probability	2
2.3	Conclusion	2
Biblio	graphy	20
3 Ra	te Adaptive Hierarchical Modulation-Assisted Two-User Opportunistic Sche	du
ing		23
3.1	Channel Model and Hierarchical Constellations	29
	3.1.1 Channel model	2
	3.1.2 Hierarchical constellations	3
		3
3.2	Two Best-User Opportunistic Scheduling	-
3.2	3.2.1 General assumptions	3
3.2	3.2.1 General assumptions 3.2.2 Overall system description	3: 3:
3.2	3.2.1 General assumptions 3.2.2 Overall system description 3.2.3 Rate adaptation process	3: 3: 3:
3.2 3.3	1wo Best-User Opportunistic Scheduling 3.2.1 General assumptions 3.2.2 Overall system description 3.2.3 Rate adaptation process Performance Analysis	3: 3: 3: 3:
3.2 3.3	1wo Best-User Opportunistic Scheduling 3.2.1 General assumptions 3.2.2 Overall system description 3.2.3 Rate adaptation process Performance Analysis	3: 3: 3: 3: 3: 3:
3.2	Wo Best-User Opportunistic Scheduling 3.2.1 General assumptions 3.2.2 Overall system description 3.2.3 Rate adaptation process 3.2.3 Rate adaptation process Performance Analysis	3: 3: 3: 3: 3: 3: 4:
3.2	3.2.1 General assumptions 3.2.2 Overall system description 3.2.3 Rate adaptation process 3.2.3 Rate adaptation process Performance Analysis	3 3 3 3 3 3 4 4 4
3.2 3.3 3.4	Wo Best-User Opportunistic Scheduling 3.2.1 General assumptions 3.2.2 Overall system description 3.2.3 Rate adaptation process 3.2.3 Rate adaptation process Performance Analysis	3: 3: 3: 3: 3: 3: 4: 4: 4: 4:
3.2 3.3 3.4 Biblio	Iwo Best-User Opportunistic Scheduling 3.2.1 General assumptions 3.2.2 Overall system description 3.2.3 Rate adaptation process 3.2.3 Rate adaptation process Performance Analysis	3: 3: 3: 3: 3: 3: 4: 4: 4: 4: 4: 4: 4: 4: 4: 4:
3.2 3.3 3.4 Biblio 4 Tw	1wo Best-User Opportunistic Scheduling 3.2.1 General assumptions 3.2.2 Overall system description 3.2.3 Rate adaptation process 3.2.3 Rate adaptation process Performance Analysis	3: 3: 3: 3: 3: 4: 4: 4: 4: 4: 4: 4:
3.2 3.3 3.4 Biblio 4 Tw les	1wo Best-User Opportunistic Scheduling 3.2.1 General assumptions 3.2.2 Overall system description 3.2.3 Rate adaptation process 3.2.3 Rate adaptation process Performance Analysis	3 3 3 3 4 4 4 4 5

V

	T. 2	Classical Single User Scheduling: Overview	52
		4.2.1 Absolute CNR-based single best user scheduling	52
		4.2.2 Normalized CNR-based single user scheduling	52
	4.3	Proposed Two-User Opportunistic Scheduling	53
•		4.3.1 Two-best user scheduling	53
		4.3.2 Newly proposed hybrid two-user scheduling	55
	4.4	Performance Analysis	56
		4.4.1 Two-best user scheduling	56
		4.4.2 Hybrid two-user scheduling	60
		4.4.3 Single user scheduling	61
	4.5	Numerical Results	63
	4.6	Conclusion	65
3i	bliogr	aphy	67
5	Que	ing Delay and Buffer Distribution of Two-User Opportunistic Schedulin	5
	Sche	mes in Wireless Networks	68
	5.1	Overall System Description and Two-User Scheduling Schemes	69
	5.1	Overall System Description and Two-User Scheduling Schemes5.1.1Rate adaptive hierarchical modulation-assisted TBS scheme	69 69
	5.1	Overall System Description and Two-User Scheduling Schemes	69 69 71
	5.1 5.2	Overall System Description and Two-User Scheduling Schemes5.1.1Rate adaptive hierarchical modulation-assisted TBS scheme5.1.2Rate adaptive hierarchical modulation-assisted HTSFormulation of the Queueing Model	69 69 71 72
	5.1 5.2	Overall System Description and Two-User Scheduling Schemes5.1.1Rate adaptive hierarchical modulation-assisted TBS scheme5.1.2Rate adaptive hierarchical modulation-assisted HTSFormulation of the Queueing Model5.2.1Markov chain analysis	69 69 71 72 72
	5.1	Overall System Description and Two-User Scheduling Schemes5.1.1Rate adaptive hierarchical modulation-assisted TBS scheme5.1.2Rate adaptive hierarchical modulation-assisted HTSFormulation of the Queueing Model5.2.1Markov chain analysis5.2.2Matrix-analytic solution for steady state distributions	69 69 71 72 72 72
	5.1	Overall System Description and Two-User Scheduling Schemes5.1.1Rate adaptive hierarchical modulation-assisted TBS scheme5.1.2Rate adaptive hierarchical modulation-assisted HTSFormulation of the Queueing Model5.2.1Markov chain analysis5.2.2Matrix-analytic solution for steady state distributions5.2.3Derivation of performance measures	69 69 71 72 72 77 77
	5.15.25.3	Overall System Description and Two-User Scheduling Schemes	69 69 71 72 72 77 77 77
	 5.1 5.2 5.3 5.4 	Overall System Description and Two-User Scheduling Schemes5.1.1Rate adaptive hierarchical modulation-assisted TBS scheme5.1.2Rate adaptive hierarchical modulation-assisted HTSFormulation of the Queueing Model5.2.1Markov chain analysis5.2.2Matrix-analytic solution for steady state distributions5.2.3Derivation of performance measuresComparison with Classical Single User SchedulingNumerical Results	 69 69 71 72 72 77 77 79 80
	5.15.25.35.4	Overall System Description and Two-User Scheduling Schemes5.1.1Rate adaptive hierarchical modulation-assisted TBS scheme5.1.2Rate adaptive hierarchical modulation-assisted HTSFormulation of the Queueing Model5.2.1Markov chain analysis5.2.2Matrix-analytic solution for steady state distributions5.2.3Derivation of performance measuresComparison with Classical Single User Scheduling5.4.1I.I.D. fading environment	 69 69 71 72 72 77 77 79 80 80
	5.15.25.35.4	Overall System Description and Two-User Scheduling Schemes5.1.1Rate adaptive hierarchical modulation-assisted TBS scheme5.1.2Rate adaptive hierarchical modulation-assisted HTSFormulation of the Queueing Model5.2.1Markov chain analysis5.2.2Matrix-analytic solution for steady state distributions5.2.3Derivation of performance measuresComparison with Classical Single User Scheduling5.4.1I.I.D. fading environment5.4.2I.N.D. fading environment	 69 69 71 72 72 77 77 79 80 80 81
	5.15.25.35.4	Overall System Description and Two-User Scheduling Schemes5.1.1Rate adaptive hierarchical modulation-assisted TBS scheme5.1.2Rate adaptive hierarchical modulation-assisted HTSFormulation of the Queueing Model5.2.1Markov chain analysis5.2.2Matrix-analytic solution for steady state distributions5.2.3Derivation of performance measuresComparison with Classical Single User SchedulingSumerical Results5.4.1I.I.D. fading environment5.4.3Simulation results for MMP arrival process	 69 69 71 72 72 77 77 79 80 80 81 85
-	 5.1 5.2 5.3 5.4 5.5 	Overall System Description and Two-User Scheduling Schemes5.1.1Rate adaptive hierarchical modulation-assisted TBS scheme5.1.2Rate adaptive hierarchical modulation-assisted HTSFormulation of the Queueing Model5.2.1Markov chain analysis5.2.2Matrix-analytic solution for steady state distributions5.2.3Derivation of performance measuresComparison with Classical Single User Scheduling5.4.1I.I.D. fading environment5.4.3Simulation results for MMP arrival processConclusion	 69 69 71 72 72 77 79 80 80 81 85 89
Bi	5.1 5.2 5.3 5.4 5.5 bliog	Overall System Description and Two-User Scheduling Schemes	69 69 71 72 72 77 79 80 80 80 81 85 89 93

6	Hier	Hierarchical Constellation for Multi-Resolution Data Transmission over Block-					
	Fadi	Fading Channels 94					
6.1 System and Channel Models			n and Channel Models				
		6.1.1	Overall system description				
		6.1.2	Hierarchical 4/16-QAM				
		6.1.3	General assumptions				
		6.1.4	Channel model				
	6.2	Propos	sed Protocol				
		6.2.1	Protocol description				
		6.2.2	Protocol modelling				
	6.3	Perfor	mance Analysis				
		6.3.1	Average packet loss rate				
		6.3.2	Average packet transmission rate				
	6.4	Nume	rical Results				
		6.4.1	Average packet loss rate				
		6.4.2	Average packet transmission rate				
	6.5	Real-t	ime Multi-Resolution Data Transmission over Correlated Fading Chan-				
		nels					
		6.5.1	Overall system description				
		6.5.2	Problem formulation				
		6.5.3	Power-distortion model				
		6.5.4	Results				
	6.6	Conclu	usion				
р	libliog	ranhv	126				
L	nnnog	гарпу					
7	Linl	k Adap	tation for OFDM Systems for Delay-Constrained Communication				
	over	r Corre	lated Fading Channels				
	7.1	Syster	n Model				
		7.1.1	Channel models				
		7.1.2	Incoming traffic model				
		7.1.3	Adaptation model				
	7.2	Sched	uling Algorithms				

.

vii

		7.2.1	Problem formulation
		7.2.2	Optimal scheduling
		7.2.3	Suboptimal algorithm
	7.3	Effect of	of Correlation among Subcarriers' Fading Gains
		7.3.1	Channel and rate allocation model
		7.3.2	Rate allocation with no delay constraints
		7.3.3	Rate allocation with hard delay constraints
,	7.4	Conclu	sions
Bil	bliogr	aphy .	
8	Vari	ance-Bo	ounded Rate Allocation for Constant Rate Media Streaming 157
	8.1	System	Model
	8.2	Variano	ce-Bounded Rate Allocation
	8.3	Consta	nt Rate Play-Out under the Receiver Buffer Outage Probability Con-
		straint	
		8.3.1	Overall description
		8.3.2	Joint PDF methodology for calculation of the receiver buffer out-
			age probability
		8.3.3	Diffusion approximation for calculation of the receiver buffer out-
	•		age probability
		8.3.4	Results
	8.4	Conclu	sions
Bi	bliogr	aphy .	
9	Con	clusion	
	9.1	Future	Research Challenges
		9.1.1	Opportunistic subcarrier allocation for OFDM-based system 176
		9.1.2	Designing multiple-best user scheduling scheme
		9.1.3	Rate adaptation for hybrid ARQ-based system for real-time multi-
		,	resolution data transmission
		9.1.4	Hybrid policy for media streaming over fading channels 177

viii

9.1.5	Effect of imperfect channel state and slow adaptation for time chan-				
	nels				
Appendices	Appendices				
Append	хА				
Append	хВ				

•

•

ix

۰.

List of Tables

List of Figures

2.1	Hierarchical 4/16-QAM constellation.	14
2.2	Normalized (by N_0) average transmit power versus the number of users	20
2.3	Markov chain representation of the buffer dynamic with power-controlled	
	modulation-assisted two-best user scheduling	22
2.4	Buffer distribution and average buffer occupancy level	23
2.5	Packet loss probability versus packet arrival probability	24
3.1	Hierarchical 4/64-QAM constellation.	30
3.2	System model of BS scheduler	33
3.3	Flow chart of proposed rate adaptive modulation-assisted TBS scheme	35
3.4	Example: Joint CNRs to constellation mapping.	38
3.5	Overall SE versus number of users (SE curves for single and two-best user	
	scheduling overlap)	44
3.6	Information access probability versus number of users	44
3.7	Average SE versus priority parameter p ($g_0 = 15[dB]$)	46
3.8	Second best user information access probability versus priority parameter	
·	$p(g_0 = 15[dB])$	47
4.1	Average system's fairness versus number of users	64
4.2	Link spectral efficiency versus number of users (link SE curves for TBS,	
	SBS and HTS overlap)	65
5.1	Buffer distribution (i.i.d. fading, Bernouli arrival with packet arrival prob-	
	ability=0.15 and number of users, K=10).	81
5.2	CDF of the queuing delay (i.i.d. fading, Bernouli arrival with packet arrival	
	probability=0.15 and number of users, K=10).	82

xi

5.3	Buffer distribution of the near-end user (i.n.d. fading, Bernouli arrival with	
	packet arrival probability=0.4 and number of users, K=10)	83
5.4	CDF of the queuing delay of the near-end user (i.n.d. fading, Bernouli	
	arrival with packet arrival probability=0.4 and number of users, K=10).	84
5.5	Buffer distribution of the far-end user (i.n.d fading, Bernouli arrival with	
	packet arrival probability=0.01 and number of users, K=10).	85
5.6	CDF of the queuing delay of the far-end user (i.n.d. fading, Bernouli arrival	
	with packet arrival probability=0.01 and number of users, $K=10$)	86
5.7	Buffer distribution (i.i.d. fading, MMP arrival and number of users, K=10).	87
5.8	CDF of the queuing delay (i.i.d. fading, MMP arrival and number of users,	
	K=10)	88
5.9	Buffer distribution of the far-end user (i.n.d. fading, MMP arrival and num-	
	ber of users, K=10)	89
5.10	CDF of the queuing delay of the far-end user (i.n.d. fading, MMP arrival	
	and number of users, K=10).	90
5.11	Buffer distribution of the near-end user (i.n.d fading, MMP arrival and	
	number of users, $K=10$).	91
5.12	CDF of the queuing delay of far-end user (i.n.d. fading, MMP arrival and	
	number of users, K=10).	92
6.1	Block diagram for multi-resolution video/image data transmission	96
6.2	Hierarchical 4/16-QAM constellation.	97
6.3	Scheduling state diagram when the maximum number of retransmissions	
	is equal to one (i.e., $N_m = 1$).	101
6.4	Scheduling state diagram when the maximum number of retransmissions	
	is equal to two (i.e., $N_m = 2$)	104
6.5	Transition out of state $S_{i,j}$ $(0 \le i \le N_m - 1, 1 \le j \le i - 1)$ when the	
	maximum number of retransmissions is equal to N_m	104
6.6	Effect of the constellation priority parameter p on the average packet loss	
	rate ($N_m = 2$ and $m = 1$)	110
6.7	Effect of the allowed maximum number of retransmissions, N_m on the av-	
	erage packet loss rate ($m = 1, p = 5, \gamma_1 = 15.39$ [dB], and $\gamma_2 = 27.69$ [dB]).	111

xii

6.8	Effect of the Nakgami m fading parameter on the average packet loss rate
	$(N_m=2, p = 5, \gamma_1 = 15.39[dB], and \gamma_2 = 27.69[dB]).$
6.9	Effect of the constellation priority parameter p on the average packet trans-
	mission rate ($N_m=2$ and $m=1$)
6.10	Effect of the Nakgami m fading parameter on the average packet transmis-
	sion rate (N_m =2, $p = 5$, $\gamma_1 = 15.39$ [dB], and $\gamma_2 = 27.69$ [dB])
6.11	Adaptive system for layered real-time image/video data transmission 117
6.12	Buffer state space
6.13	Comparison of hierarchical and uniform constellation-based schemes 124
7.1	Scheme of the OFDM system
7.2	Optimal delay and aggregate BER trade-offs with different number of car-
	riers (Doppler frequency, $f_d = 300$ Hz)
7.3	Optimal delay and power tradeoffs with different number of carriers (Doppler
	frequency, $f_d = 100$ Hz)
7.4	Optimal and sub-optimal delay and aggregate BER trade-offs for different
	fading rates in two carrier case
7.5	Optimal and sub-optimal delay and aggregate BER trade-offs for different
	fading rates in three carrier case
7.6	Sub-optimal delay and aggregate BER trade-offs for different fading rates
	in four carrier case
7.7	Optimal and sub-optimal delay and power trade-offs for different fading
	rates in two carrier case.
7.8	Optimal and sub-optimal delay and power trade-offs for different fading
	rates in three carrier case
7.9	Sub-optimal delay and power trade-offs for different fading rates in four
	carrier case
7.10	Power versus rate constraint for delay limited and non-delay limited rate
	allocation
8.1	Scheme of the variance-bounded rate allocation
8.2	Channel gain versus transmission rate ($\bar{R} = 2$ [Mbits/Sec], $W = 1$ [MHz]
	and $h_0 = 10$)

,

.

	providence and the second s
8.3	Variance versus transmit power $(h_0 = 10)$
8.4	Buffer outage probability versus transmission rate variance ($W = -1$ [MHz],
	and $h_0 = 10$)
8.5	Buffer outage probability versus transmit power ($W = 1$ [MHz], and $h_0 =$
	10)

1

xiv

List of Abbreviations

ADSL	:	Asynchronous digital subscriber line
ARQ	:	Automatic repeat request
ASS	:	Absolute CNR-based single user scheduling
AWGN	:	Additive white Gaussian noise
BER	:	Bit error rate
BS	:	Base station
CDF	:	Cumulative distribution function
CDMA	:	Code division multiple access
CMDP	:	Constrained Markov decision process
CNR	:	Carrier-to-noise ratio
CRC	:	Cyclic redundancy check
CSI	:	Channel state information
DA	:	Diffusion approximation
DCT	:	Discrete cosine transform
DOF	:	Degree of fairness
I.n.d.	:	Independent but non-identical distribution
DTMC	:	Discrete time Markov chain
DVB-T	:	Digital video broadcasting-Terrestrial
FEC	:	Forward error correction
Fps	:	Frame per second
FSMC	:	Finite state Markov chain
GA	:	Gaussian approximation
HTS	:	Hybrid two-user scheduling
I	:	In-phase
I.i.d.	:	Identically and independently distributed
I.n.d.	:	Independently but non-identically distributed

ISI	:	Inter symbol interference
JPDF	:	Joint probability density function
LP	:	Linear Programming
LSB	:	Least significant bit
LSP	:	Least significant part
MC	:	Multicode
MCM	:	Multicarrier modulation
MDP	:	Markov decision process
MIMO	:	Multi-input multi-output
MMP	:	Markov modulated Poisson
MSB	:	Most significant bit
MSP	:	Most significant part
NSS	:	Normalized CNR-based single use scheduling
OFDM	:	Orthogonal frequency division multiplexing
OFDMA	:	Orthogonal frequency division multiplexing access
PDF	:	Probability density function
PER	:	Packet error rate
PFS	:	Proportional fair scheduling
PI	:	Policy iteration
PLR	:	Packet loss rate
PSK	:	Phase shift keying
Q	:	Quadrature phase
QAM	:	Quadrature amplitude modulation
QBD	:	Quasi birth-death
QoS	:	Quality of Service
RVI	:	Relative value
RW	:	Random walk
SBS	:	Single best user scheduling
SE	:	Spectral Efficiency
TBS	:	Two-best user scheduling
TD	:	Temporal diversity
TDMA	:	Time division multiple access

xvi

UMDP : Unconstrained Markov decision process

VBRA : Variance-bounded rate adaptation

1

,

xvii

Acknowledgements

I would first like to express my gratitude towards my supervisor, Professor Vijay K. Bhargava without whose guidance, attention to detail, and encouragement, I could not have completed this work. At the same time, special thanks go to Prof. Mohamed-Slim Alouini for his invaluable suggestions and helpful advice. The friendly and supportive atmosphere inherent to the Communication Group at UBC contributed essentially to the final outcome of my studies. I would like to take this opportunity to thank some of my fellows Dr. Dejan, Mamun, and Chandika.

It is not possible to mention all the people that have in some way influenced this work. Specially, my wife Sharmin Ferdous helped me a lot while I was warping up my thesis through her continuous support. Last, but not the least, the biggest personal thanks goes to my family, especially to my parents.

Dedication

To my beloved parents and to the members of my family whose blessing, love and affection have been the greatest possession in my life.

Statement of Co-Authorship

I was the primary researcher for this dissertation. The details of my contributions are as follows. For contributions is Chapters 2, 3, 4, 5, 6 and 7, I have identified and formulated the research problems. For the contribution presented in Chapter 8, my co-author has partly identified and formulated the research problem. I have analyzed performances for all of our designed schemes. These analyzes have resulted closed-form expressions for useful performance metrics. I have also written MATLAB program and evaluated the performance via computer simulation. I have compared computer simulation results with analytical counterparts. Finally, I have materialized the ideas and prepared manuscripts for scholarly publication.

.....

Chapter 1

Introduction

Wireless networks are expected to support a wide variety of services. Some examples of these services are video conferencing, Internet access, email transfer, and web-browsing. Different applications have different traffic characteristics and quality of service (QoS) requirements. These QoS parameters include delay constraint, packet loss rate (PLR) and packet error rate (PER) requirements. For example, depending on the throughput or delay requirements of the services, wireless multimedia services have already been classified as guaranteed or best-effort services [1]. In guaranteed services, the guarantee may be in the form of ensuring that the throughput is greater than some minimum value, or that the delay experienced is smaller than some threshold. Mobile wireless links are time-varying in nature and error prone. In addition, communication resources, for example, transmit power and bandwidth, are scarce. Through a cross-layer design approach, these inherent quality parameters of different services can facilitate efficient use of communication resources (see, for example, [2]). Specifically, the relative delay tolerance of data applications, with the bursty activity patterns and time-varying nature of wireless channels, opens up the possibility of scheduling transmissions to efficiently allocate limited wireless communication resources.

Common thread of our research presented in this dissertation is to design innovative schedulers by joint consideration of the time-varying nature of wireless channels as well as the QoS parameters of the services that are to be transmitted. Specifically, we design schedulers for two-user opportunistic scheduling, multi-resolution image/video data transmission, rate adaptation for high-quality media streaming, and link adaptation for orthogonal frequency division multiplexing (OFDM)-based systems for delay-constrained communications. The design goals of these schedulers depend on the specific applications for which they are designed. For example, in the case of two-user opportunistic scheduling, the base station (BS) scheduler selects the user(s) as well as the number of packets that are to be transmitted to the selected user(s). On the other hand, the scheduler for multi-

resolution image/video data transmission selects the packets that are to be transmitted for a given channel condition and a power budget in order to maintain the quality of the received video/image. In what follows, we provide a general background and motivation for the specific research work reported in this dissertation.

1.1 General Background and Motivation

1.1.1 Opportunistic scheduling

Independent channel variations among the users in a multi-user environment can be exploited to increase the system's overall throughput by opportunistic scheduling (also known as multi-user diversity) [3], [4]. In opportunistic scheduling, the scheduler at the BS selects a single "best" user in each time slot based on available channel state information, and transmits information only to this selected user. Although this classical single best user scheduling (SBS) maximizes the average overall throughput, it suffers from one main disadvantage, in that the serviced users may have a low frequency of information access. The low frequency of information access has two end effects. First, it leads to a certain degradation of system fairness. In a heterogenous fading environment where different users have different average link gains, the users with relatively strong average link gains enjoy a higher frequency of information access than the users with weaker average link gains; thus, long-term fairness degrades. Second, in a homogenous environment where the users have the same average link gain, the low information access frequency with SBS can result in higher queuing delay of the packets at the transmitter buffer. For a memory-limited system, the PLR can be quite high. In this regard, let us consider a system of K users with identical average link gains, and with each user having a transmission buffer at the transmitter where packets are queued until they are served to the desired users. We assume a fixed transmission rate of K packets per transmission slot. With SBS in this homogenous environment, on average each user will get access to one out of K slots, and will receive K packets. Therefore, the average waiting time for the packets at the transmission buffer is almost Ktime slots. On the other hand, if the frequency of information is doubled, on average each user will get access to one out of K/2 slots, and will receive K/2 packets. Although in both cases the users will have same the average throughput, the latter one will offer less queuing

delay of the packets at the transmission buffer. In addition, the delay jitter (variance of the delay) will be reduced with higher frequency of channel access as higher frequency of information access smoothes out the packet delivery process to the users. The increased queuing delay can also result considerable packet loss due to buffer over-flow for a system with limited buffer size.

In a wireless network where different users are located in different locations from the BS resource sharing among the users may be monopolized by the users who are located close to the BS. Consequently, fairness among the users decreases [4]. In order to alleviate the fairness issue, proportional fair scheduling (PFS) was proposed in [4]. The PFS achieves better fairness among the users at the expense of a certain amount of system spectral efficiency (SE). Therefore, there is a trade-off between achievable multiuser diversity gain and fairness among the serviced users.

The above-mentioned constraints motivate the design of an opportunistic scheduling scheme that can offer more frequent channel access yet still exploit the multiuser diversity gain. Recently, the concept of multiple best user scheduling has been proposed for multicode (MC) code division multiple access (CDMA) systems [5], [6]. The schemes proposed in [5] and [6] transmit information to two users by assigning a different number of orthogonal codes to these users thereby improving system fairness. A generalized opportunistic scheduling scheme for multiple users simultaneously using multiple interfaces, for example, multicarrier CDMA, multi-input and multi-output (MIMO) and orthogonal frequency division multiple access (OFDMA) systems, has been studied in [7]. All of these schemes consider that there are a number of parallel transmission channels between the users and the access point/BS. Contrary to the multiple interface-based multiple best user scheduling schemes examined in [5], [6], [7], in this dissertation, we propose modulation-assisted two-user opportunistic scheduling schemes. Our work differs from the existing schemes in that it proposes a modulation-assisted solution and considers that there is only one transmission link between the users and the access point/BS. Our proposed scheme can be used in a narrow-band time division multiple access (TDMA) system as well.

3

1.1.2 Multi-resolution image/video data transmission

As we mentioned earlier, the present and next generation of mobile communication systems will support the transmission of images as well as video. Some examples of these types of transmission are video conferencing, emergency medical services, and site surveys. For reliable digital data transmission over these error prone channels, there are mainly two error-control strategies, namely forward error correction (FEC) schemes and automatic repeat request (ARQ) schemes [8]. Ideally, with ARQ protocols, an error-free transmission is possible if the allowed maximum number of retransmissions is equal to infinity. In real-time applications, such as in video transmission, an arbitrary large number of retransmissions for certain packets may lead to large round-trip delays and buffer sizes. Therefore, ARO schemes with a limited number of maximum retransmissions (also known as truncated, constrained retransmission ARQ), are typically used in many applications (see, for example, [9]). This class of ARQ schemes meets a prescribed delay requirement of services by dropping an erroneous packet at the data link layer after a certain number of retransmissions. All ARQ schemes require redundancy, and involve decoding complexity as well as delay for using error-detecting codes. Recently, a carrier-to-noise ratio (CNR) threshold-based¹ retransmission strategy was proposed in [10] in order to reduce redundancy, decoding complexity, and delay. The proposed system assumed that the service (e.g., video-to-car) to be transmitted has a target PER, PER₀, as well as a delay constraint. Therefore, a packet is considered to be received "unsuccessfully" and a retransmission is requested if the instantaneous CNR is less than a CNR threshold (determined by PER₀) during the transmission of the packet.

On the other hand, most of video streams have a layered structure where different layers have different transmission priorities and sensitivities. For instance, MPEG-4 embraces the multi-resolution model as a mechanism to spatially and temporally adjust a video stream to a given channel bandwidth. To transmit these different multi-resolution (i.e., layered) video streams, rather than applying a uniform data protection method to the whole video stream, it is more efficient if the higher sensitive portion of video data is protected by a higher reliable data protection [9] method. Similarly, some source coding schemes, such as

4

¹Throughout this dissertation, CNR threshold means the required CNR value in order to maintain a target bit error rate (BER).

discrete cosine transform (DCT), can divide images into several parts depending on their importance. In these source coders, the sensitivity of the channel error varies significantly depending on the importance of the corresponding bits. In the context of multi-resolution transmissions, video streams or image sources can be encoded into classes of different resolution levels with the following main features.

- 1. Higher-level information cannot be used unless all of the lower-level information is received correctly. For instance, if Level 1 represents the lowest Level of resolution and Level 2 represents the successive higher level of information, Level 2 information is useless if Level 1 information is not received correctly.
- 2. Since higher-level information is useless without all the lower-level information, different resolution levels have different transmission priorities. Specifically, a given resolution level has a higher transmission priority than its successive higher resolution levels. For example, Level 1 information has a higher transmission priority than Level 2 information.

A hierarchical image transmission scheme using a hierarchical quadrature amplitude modulation (QAM) constellation was proposed in [11] to protect different resolution components differently. This scheme proposed to devote the first level of hierarchy (i.e., the highest protected sub-channels) of the hierarchical 4/16-QAM constellation for basic information transmission while transmitting refinement information in the second level of hierarchy (i.e., the lowest protected sub-channels). Non-uniform signal constellations in conjunction with the ARQ scheme were used in [12] to adapt the data rate of a single resolution message string transmitting over fading channels. For the same reason as [10], the system proposed in [12] also used an instantaneous CNR threshold-based retransmission scheme. Inspired by this application of non-uniform constellations, in Chapter 6 of this dissertation, we propose a new multi-resolution data transmission technique over block-fading channels.

1.1.3 Link adaptation for OFDM-based systems

OFDM, a form of multicarrier modulation scheme [15], is a promising technology designed to meet the demands of high data rate traffic for wide band wireless communications. In an OFDM system, a high data rate system is divided into a number of parallel low data rate streams, each of which is then modulated in a given narrow band frequency channel using orthogonal subcarriers. Each narrow band channel (also called a sub-carrier) undergoes frequency flat fading and thus eliminates the complex equalizer design at the receiver in order to combat intersymbol interference (ISI) in multi-path fading environments. Another advantage of the OFDM modulation technique is the possibility of exploiting different fading gains among the subcarriers in a given channel access and thus loading (power or bit or both power and bit) different subcarriers according to their respective gains. Due to its high spectral efficiency and resistance to multi-path fading channels, the OFDM modulation technique is already being used in several fixed and mobile radio systems, such as asymmetric digital subscriber lines (ADSL), and wireless LANs (IEEE 802.11a/g and IEEE 802.16). It has also been standardized for digital video terrestrial (DVB-T) and audio broadcasting [16].

Since different subcarriers in an OFDM system may have different fading gains in a given channel access, use of the same modulation order in all subcarriers leads to inefficient utilization of the overall spectrum [17]. Assuming that the channel state information (CSI) for all subcarriers is available at the transmitter, different power or bit or both power and bit loading schemes have been proposed in literature. Different loading algorithms have different end goals [18]. One broad class of bit-loading algorithms minimizes the transmit power while attaining a fixed transmission rate as well as a given target BER (see, for example, [19], [20]). In another version of bit-loading algorithms, average capacity is maximized at a fixed transmit power. These are the two extreme of link adaptation algorithms from a delay perspective. The loading algorithms that maximize average capacity or transmission rate are suitable for services that have no delay constraint, theoretically infinite delay. On the other hand, the algorithms that maintain a fixed transmission irrespective of channel quality are useful for services that have hard delay constraints.

In order to allocate communication resources efficiently, recently adaptive transmission techniques, which vary transmission rate or power jointly considering channel quality, traffic arrival statistics and the buffer state have been proposed in [21], [22]. Through crosslayer design and optimization, the proposed schemes study the optimal trade-offs between different conflicting objectives, for example, buffering delay and transmit power, in single carrier transmission systems. In Chapter 7 of this dissertation, we develop link adaptation

6

algorithms for a multicarrier system for generalized delay-constrained communications.

1.1.4 Variance-bounded rate adaptation for delay-constrained communications

In order to take advantage of the time-varying nature of wireless channels, transmit power, rate or both rate and power are varied according to the channel quality [13]. Specifically, for a fading wireless channel, the water-filling policy has been proven to maximize average transmission capacity for an average transmit power constraint [13]. The water-filling policy suggests transmitting information at a higher rate with higher power when the channel quality is relatively better. As a consequence, the packets may need to wait at the transmitter buffer until the channel quality becomes relatively better. However, in order to satisfy some upper-layer QoS parameters, e.g., delay requirements, buffer underflow probability, the transmission system should not only maintain an average transmission rate but also may require guaranteeing a bound on the variance of the transmission rate. For example, in case of high-quality video streaming over wireless links, there should sufficient packets at the play-out buffer so that the video frames can be played in a timely manner without interruption (see, for example, [14]). In other words, in order to play out the entire video sequence smoothly, the buffer should not experience an underflow. Constant rate downloading from the server can avoid buffer underflow. But this constant rate downloading policy may not lead to a power-efficient downloading scheme. On the other hand, from channel quality variation point of view, packet transmission should be scheduled at a higher rate during relatively better channel qualities, as suggested by the water-filling algorithm. This waterfilling policy, which guarantees an average rate constraint, may cause too much play-out buffer outage due to the unavailability of the video frames on time. Since the video frames are to be available at the buffer at their taken out time instants, instead of having constant rate download some variance in the downloading rate can be tolerated.

Therefore, a question arises as to how to schedule transmissions such that minimize the transmit power is minimized while the play-out buffer outage is kept very low, i.e. below a desired level. Motivated by the aforementioned question, in Chapter 8, we propose a new rate allocation scheme, namely *variance-bounded rate adaptation (VBRA)*. Our newly proposed VBRA scheme is not limited to "on-demand" media streaming. In general, it is

useful to any application where the transmission rate is bounded by a given variance that is related to some upper layer QoS parameters, e.g., delay, jitter. A closed-form solution of the optimal transmission rate, which minimizes the transmit power while maintaining a given average transmission rate and a variance of the transmission rate, is presented. Using this solution, we explore the optimal power and variance trade-offs for a given average rate constraint.

1.2 Thesis Outline

The remainder of this dissertation is organized as follows. Chapter 2 presents a two-best user opportunistic scheduling (TBS) scheme using power-controlled hierarchical modulation and Chapter 3 details a rate adaptive hierarchical modulation-assisted TBS scheme. The performances of these schemes are also analyzed in respective chapters. In Chapter 4, we analyze the performance of the TBS in a heterogenous wireless network where different users have different average link gains. In this chapter, we also propose a hybrid two-user scheduling (HTS) scheme. In Chapter 5, we analyze delay statistics and buffer distribution of the TBS scheme using queuing analysis. In Chapter 6, we propose adaptive scheduling protocols for multi-resolution image/video data transmission over fading channels. Chapter 7 presents transmission frameworks for an OFDM-based system for generalized delay-constrained communications. In Chapter 8, we describe VBRA scheme for media streaming over time wireless links. Finally, Chapter 9 concludes the dissertation and outlines some future interesting research avenues ensuing it.

Bibliography

- [1] S. Shakkottai, T. S. Rappaport, and P. C. Karlsson, "Cross-layer design for wireless networks," *submitted for possible Journal publication*, Jun. 2003, also available online: *http://www.ece.utexas.edu/ shakkott/Pubs/cross-layer.pdf*.
- [2] R. Berry and R. Gallager, "Communication over fading channels with delay constraints," *IEEE Trans. Inform. Theory*, vol. 48, pp.1135-1149, May 2002.
- [3] R. Knopp and P. A. Humblet "Information capacity and power control in single-cell multiuser communications," in *Proc. IEEE Int. Conf. Commun. (ICC'95)*, Seattle, USA, Jun. 1995, pp. 331-335.
- [4] D. N. Tse, P. Viswanath, and R. Laroia, "Opportunistic beamforming using dumb antennas," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1277-1294, Jun. 2002.
- [5] D. I. Kim, "Two-best user scheduling for high-speed downlink multicode CDMA with code constraint,", in *Proc. IEEE Conf. Global Telecommun. (Globecomm' 04)*, Dallas, TX, Oct. 2004, pp. 2659- 2663.
- [6] S. Dost, M. Oguz Sunay, and V. K. Bhargava, "A feasibility study of two user downlink transmission for IS-856 system,", in *Proc. Personal and Indoor Commun. Conf.* (*PIMRC'04*),, Bercelona, Spain, Sep. 2004, pp. 2029 - 2033.
- [7] S. Kulkarni and C. Rosenberg, "Opportunistic scheduling for wireless systems with multiple interfaces and multiple constraints," in *Proc. of the 6th ACM international* workshop on Modeling analysis and simulation of wireless and mobile systems, San Diego, CA, USA, Sep. 2003, pp. 11-19.
- [8] S. Lin and D. J. Costello, Jr., Error Control Coding: Fundamentals and Applications, Prentice Hall: Englewood Cliffs, NJ, 1983.

- [9] C.-C. Liu and S.-C. S. Chen, "Providing unequal reliability for transmitting layered video streams over wireless networks by multi-ARQ schemes," in *Proc. IEEE Int. Conf. Image Processing (ICIP' 99)*, Kobe, Japan, Oct. 1999, pp. 100-104.
- [10] Q. Liu, S. Zhou, and B. Giannakis, "Jointly adaptive modulation and packet retransmission over block-fading channels with robustness to feedback latency," in *Proc. Conf. Information Sciences and System (CISS 2003)*, The Johns Hopkins University, USA, Mar. 2003.
- [11] M. Morimoto, M. Okada, and S. Komaki, "A hierarchical image transmission system in fading channel," in *Proc. IEEE Int. Conf. Univ. Personal Commun. (ICUPC' 95)*, Tokyo, Japan, Oct. 1995, pp. 769-772.
- [12] R. Otnes and T. Maseng, "Adaptive data rate using ARQ and non-uniform constellations," in *Proc. IEEE Veh. Technol. Conf. (VTC' 01-Spring)*, Rhodes, Greece, May 2001, pp. 1211-1215.
- [13] A. J. Goldsmith, "The capacity of downlink fading channels using variable power and variable rate," *IEEE Trans. Veh. Technol.*, vol. 46, pp. 569-580, Aug. 1997.
- [14] Y. Li, A. Markopoulou, N. Bambos, and A. J. Apostolopoulos, "Joint power/playout control schemes for media streaming over wireless links," *IEEE Trans. Multimedia*, vol. 8, pp. 830-843, Aug. 2006.
- [15] J. A. C. Bingham, "Multicarrier modulation for data transmission: an idea whose time has come," *IEEE Commun. Mag.*, vol. 28, pp. 5-14, May 1990.
- [16] DVB-T Standard: ETS 300 744, Digital Broadcasting Systems for Television, Sound and Data Services: Framing Structure, Channel Coding and Modulation for Digital Terrestrial Television, ETSI Draft EN300 744, 1999-1, 1999.
- [17] T. Keller and L. Hanzo, 'Multicarrier modulation: a convenient framework for timefrequency processing in wireless communications," *Proc. of the IEEE*, vol. 88, pp. 611-640, May 2000.

- [18] A. T. Toyserkani, J. Ayan, S. Naik, Y. Made, and O. Al-Askary, 'Sub-carrier based adaptive modulation in HIPERLAN/2 system," in *Proc. IEEE Int. Conf. Commun.* (*ICC'04*), Paris, France, Jun. 2004 pp. 3460-3464.
- [19] L. Goldfeld and V. Lyandres, 'Capacity of the multicarrier channel with frequencyselective Nakagami fading," *IEICE Trans. Commun.*, vol. E83-B, pp. 697-702, Mar. 2000.
- [20] C. Y. Wong, R. S. Cheng, K. B. Letaief, and R. D. Murch, "Multiuser OFDM with adaptive subcarrier, bit, power, and power allocation," *IEEE J. Select. Areas Commun.*, vol. 17, no. 10, pp. 1747-1758, Oct. 1999.
- [21] A. T. Hoang and M. Motani, "Buffer and channel adaptive modulation for transmission over fading channels," in *Proc. IEEE Int. Conf. Commun. (ICC'03)*, Anchorage, AK, vol. 4. pp. 2748-2752, May 2003.
- [22] D. Djonin, A. K. Karmokar, and V. K. Bhargava, "Delay-constrained rate and power adaptation over correlated fading channels," in *Proc. IEEE Global Teleommun. Conf.* (*Globecom'04*), Dallas, TX, Sep. 2004, pp. 3448- 3453.

11

Chapter 2

Multi-User Opportunistic Scheduling Using Power-Controlled Hierarchical Constellations

²As it was mentioned in Section 1.1.1 that classical SBS scheme provides low frequency of information access to the serviced users in a wireless network. This leads to higher average queuing delay and packet loss rate for a finite buffer system. On the other hand, hierarchical constellations consist of non-uniformly spaced signal points (see, for example, [23], [24]). Due to the capability of providing different levels of protection, this type of constellations has been proposed in many applications including broadcast channel [27], multimedia transmission [25], downlink multiplexing [26] and superposing bits from different users in the same sub-carrier of an OFDM system [31]. It has also been included in digital video broadcasting (DVB-T) [28], MediaFLO [29], Ultra Mobile Broadband (UMB), a 3.5th generation mobile network standard developed by 3GPP2 [30].

The above mentioned power efficient hierarchical multiplexing schemes motivates us to study an application of hierarchical constellations in *multi-user* opportunistic scheduling which transmits information to multiple best users simultaneously rather than a single best user in a given time slot. This way the users are given more frequent information access in order to reduce queuing delay and loss rate of the packets at the transmission buffer. The detailed description of our proposed scheme will be given in Section 2.1.2. We present expressions for the average transmit power, and the outage probability with truncated channel inversion power control for a fading environment where the users' channels' fading gains are identically and independently distributed (i.i.d). Using these expressions,

²The research work presented in this chapter has been published as Md. J. Hossain, M.-S. Alouini, and V. K. Bhargava, "Multi-user Opportunistic Scheduling using power-controlled Hierarchical Constellations", *IEEE Trans. Wireless Commun.*, Vol. 6, pp. 1581-1586, May 2007.

we explore the trade-offs between these higher layer performances and the power efficiency of multi-user opportunistic scheduling using hierarchical constellations. We also analyze higher layer performances, for example, buffer distribution, average buffer occupancy, and loss probability of the packets at the transmission buffer via queuing analysis. We finally compare in this chapter the hierarchical multi-user scheduling scheme with a time-slotted multi-user opportunistic scheduling. In this time-slotted scheme two or more best users are scheduled for transmission in a mini time-slotted fashion.

2.1 Multi-User Opportunistic Scheduling

2.1.1 Hierarchical constellations

Throughout the chapter, we limit ourselves *two best user* opportunistic scheduling using hierarchical 4/16-QAM constellation which is constructed with two levels of hierarchy. However, our approach can easily be generalized for the scheduling of more than two best users using higher order hierarchical constellations [24].

A hierarchical 4/16-QAM constellation (see, for example, [23]), is shown in Fig. 2.1 and can be modelled as follows. We assume that there are two streams of data where one stream requires higher level of protection than the other. For every channel access two bits are chosen from each level. The two bits requiring the highest protection are assigned to the most significant bit (MSB) position in the inphase (I) and quadrature (Q) phase channels. Consequently, the two bits that require the least protection are assigned to the least significant bit (LSB) position in the I and Q channels. In Fig. 2.1, the fictitious black symbols represent a 4-QAM constellation (referred to as the first hierarchy and denoted by H_1). The distance between the symbols in the first hierarchy is represented by d_1 . The actual transmitted symbols are the white symbols and they represent a 16-QAM constellation. This is the second level of hierarchy (denoted by hierarchy H_2). One of these white symbols surrounding a selected black symbol in the first hierarchy is selected by the two bits that requires the least protection. The distance between the symbols in the second hierarchy is denoted by d_2 .

The exact BER expressions over additive white Gaussian noise (AWGN) channel for this hierarchical 4/16-QAM constellation were derived in [24]. The BER equations in [24]



Figure 2.1: Hierarchical 4/16-QAM constellation.

are not invertible in terms of transmit power and the constellation priority parameter for given other parameters. However, the well known SNR-gap approximation [33] can be easily inverted in terms of the transmit power and this leads to a simple design of hierarchical constellations (see [31]). As such, we use this SNR-gap approximation to evaluate the average transmit power analytically. According to the SNR-gap approximation, the average power of kth hierarchy is given by [31], [33]

$$\mathbf{P}_k = \frac{d_k^2 (2^m - 1)}{6} \tag{2.1}$$

where d_k is the distance between the constellation symbols in the kth hierarchy and m is the number of bits assigned to kth hierarchy which is equal to two in the case of hierarchical 4/16-QAM. The distances d_1 and d_2 as shown in Fig. 2.1 are calculated recursively as follows. The distance between the actual constellation (16-QAM) symbols, d_2 is calculated as

$$d_2 = \sqrt{6 \times \frac{2N_0\Theta}{\alpha^2}},\tag{2.2}$$

where α^2 is the received power channel gain of the user assigned to second hierarchy, N_0 is the one sided noise power spectral density, and Θ is the SNR-gap of the uncoded QAM constellation and given by

$$\Theta = \frac{1}{3} \left[Q^{-1} \left(\frac{\text{BER}_0}{4} \right) \right]^2, \qquad (2.3)$$

where $Q^{-1}(\cdot)$ is the inverse standard Gaussian Q-function and BER₀ is the target BER. The distance between the actual symbols in different quadrant d' as shown in Fig. 2.1, can be calculated as

$$d' = \sqrt{6 \times \frac{2N_0\Theta}{\beta^2}},\tag{2.4}$$

where β^2 is the received channel fading power gain of the user assigned to the first hierarchy. Now the distance between fictitious symbols (4-QAM) in the first hierarchy can be written as (see Fig. 2.1)

$$d_1 = d' + d_2. (2.5)$$

Using Eqs. (2.1)-(2.5), the total transmit power with hierarchical 4/16-QAM constellation (assuming that the second hierarchy is assigned to a user with channel fading gain α^2 and that the first hierarchy is assigned to a user with channel fading gain β^2) can be expressed

as

$$\mathbf{P}(\alpha,\beta) = \frac{C_2}{\alpha^2} + \frac{C_2}{\alpha^2 \beta^2} (\alpha^2 + 2\alpha\beta + \beta^2)$$
(2.6)

where the constant $C_2 = 2N_0\Theta(2^2 - 1)$.

2.1.2 Hierarchical constellation-based multi-user opportunistic scheduling

To describe our proposed multi-user opportunistic scheduling let us consider a BS transmitting to K users opportunistically. The users have to meet a predetermined target BER, BER₀ and the channel variation among the users is assumed to be i.i.d. The channel fading amplitude of a given user is Rayleigh faded which corresponds to the following probability density function (PDF) for the kth user channel fading gain h_k , $p_{h_k}(h_k)$

$$p_{h_k}(h_k) = \frac{1}{h_0} \exp\left(-\frac{h}{h_0}\right); \quad h_k \ge 0$$
 (2.7)

where h_0 is the identical average channel fading gain of the users.

We now describe the scheduling as well as the corresponding adaptation process.

- Step 1: Based on users' channel qualities, the scheduler at the BS ranks the users and picks up the first (denoted by U₍₁₎) and the second best user (denoted by U₍₂₎). Without loss generality, let us consider the fading gains of the users U₍₁₎ and U₍₂₎ are denoted by h₍₁₎ and h₍₂₎, respectively.
- Step 2: If the channel fading power gain of the second best user, $h_{(2)}$ is above a threshold h_{μ}^2 , go to the next step. Otherwise declare a transmission outage for that time step and restart from step 1 in the next time slot.
- Step 3: Since the user, $U_{(1)}$ requires less protection than user $U_{(2)}$, hierarchy H_2 is assigned to the user $U_{(1)}$ whereas hierarchy H_1 is assigned to the user $U_{(2)}$. With this assignment, the transmitter adapts the transmission power as well as the relative distances between the symbols $(d_1, d_2 \text{ and } d')$ such as the target BER, BER₀ is met with the corresponding channel fading power gains $h_{(1)}$ and $h_{(2)}$.

The position of the bits in a transmitted symbol for the selected users are sent via a feedforward channel. Thus the selected users receive the symbols and look only for the bits
in particular positions within the symbol. Since the average transmit power with channel inversion power control policy on Rayleigh faded channel is infinite [32], a truncation threshold h_{th}^2 is used in the channel inversion power control. This threshold h_{th}^2 is set to a value such that a pre-specified transmission outage probability is met. The expression for the outage probability will be given in closed-form in Section 2.2.1. In Step 3, the required power as well as the constellation parameter can be found numerically solving BER expressions of the hierarchical 4/16-QAM [24]. However, as mentioned above for analytical tractability, we use the approximations in Eq. (2.6).

2.1.3 Comparison with uniform QAM constellation-based scheme

Multi-user scheduling is also possible using uniform QAM modulation in a time division fashion. Basically, the whole transmission slot is divided into a number of mini-slots and in each mini-slot the data of a particular user is transmitted. In two user scheduling, the transmission slot is divided into two mini-slots. The first best user is scheduled in the first mini-slot whereas the second best user is scheduled in the second mini-slot. The transmit power in each mini-slot is adjusted according to the corresponding selected user's channel gain so that the bits are transmitted below the target BER using uniform 16-QAM over an AWGN channel.

2.2 Performance Analysis and Results

In this section, we derive analytical expressions for the average transmission power and the outage probability. We also analyze the buffer distribution function, the average buffer occupancy, and the PLP for all the schemes under consideration. Some selected numerical results obtained via computer simulations are also presented to compare the performance of the schemes under consideration. We use an uncoded target BER, BER₀ of 10^{-4} and a pre-specified outage probability of 10^{-3} . We assume channel variation among the users is i.i.d with an average channel fading gain of 5[dB].

2.2.1 Outage probability

For two-user opportunistic scheduling a transmission outage occurs if the channel fading gain of the second best user, $h_{(2)}$ is less than a threshold h_{th}^2 . Therefore, the outage probability of two-user scheduling, O_{out}^2 can be written as

$$O_{\text{out}}^2 = \int_0^{h_{\text{th}}^2} \int_{h_{(2)}}^{+\infty} p_{h_{(1)}h_{(2)}}(h_{(1)}, h_{(2)}) dh_{(1)} dh_{(2)}$$
(2.8)

where $p_{h_{(1)}h_{(2)}}(h_{(1)}, h_{(2)})$ is the joint PDF of the largest two channel gains and can be expressed as [34]

$$p_{h_{(1)}h_{(2)}}(h_{(1)}, h_{(2)}) = \frac{K(K-1)}{(h_0)^2} \exp\left(-\frac{h_{(1)}}{h_0}\right) \exp\left(-\frac{h_{(2)}}{h_0}\right) \\ \times \left(1 - \exp\left(-\frac{h_{(2)}}{h_0}\right)\right)^{K-2}; h_{(1)} \ge h_{(2)} \ge 0.$$
(2.9)

Using Eq. (2.9) we can evaluate the integral in Eq. (2.8) as follows:

$$O_{out}^{2} = \int_{0}^{h_{th}^{2}} \int_{h_{(2)}}^{+\infty} \frac{K(K-1)}{(h_{0})^{2}} \exp\left(-\frac{h_{(1)}}{h_{0}}\right) \exp\left(-\frac{h_{(2)}}{h_{0}}\right) \\ \times \left(1 - \exp\left(-\frac{h_{(2)}}{h_{0}}\right)\right)^{K-2} dh_{(1)} dh_{(2)} \\ = 1 - \sum_{i=0}^{K-2} {\binom{K-2}{i}} (-1)^{i} \frac{K(K-1)}{i+2} \exp\left(-\frac{(i+2)h_{th}^{2}}{h_{0}}\right).$$
(2.10)

For single-user scheduling a transmission outage occurs if the channel gain of the first best user is below a threshold h_{th}^1 and thus the outage probability of single-user scheduling, O_{out}^1 can be written as

$$O_{out}^{1} = \int_{0}^{h_{th}^{1}} p_{h_{(1)}}(h_{(1)}) dh_{(1)}$$
(2.11)

where $p_{h_{(1)}}(h_{(1)})$ is the PDF of the largest channel gain and is known to be given by [34]

$$p_{h_{(1)}}(h_{(1)}) = \frac{K}{h_0} \exp\left(-\frac{h_{(1)}}{h_0}\right) \left(1 - \exp\left(-\frac{h_{(1)}}{h_0}\right)\right)^{K-1}.$$
 (2.12)

Using Eq. (2.12) in Eq. (2.11), the outage probability for single user scheduling can be written in the desired closed-form

$$O_{out}^{1} = 1 - \sum_{i=0}^{K-1} {\binom{K-1}{i}} (-1)^{i} \frac{K}{i+1} \exp\left(-\frac{(i+1)h_{th}^{1}}{h_{0}}\right).$$
(2.13)

Solving Eqs. (2.10) and (2.13) for a given transmission outage, we can find the values of the thresholds h_{th}^2 and h_{th}^1 , respectively.

2.2.2 Average transmit power

Averaging Eq. (2.6) over the channel gain joint PDF of the two best users, the average transmit power in two-user hierarchical constellation-based scheduling, $P_{\rm H}^2$ can be expressed as

$$\mathbf{P}_{\mathbf{H}}^{2} = \int_{h_{\text{th}}^{2}}^{+\infty} \int_{h_{(2)}}^{+\infty} \left[\frac{C_{2}}{h_{(1)}} + \frac{C_{2}}{h_{(1)}h_{(2)}} (h_{(1)} + 2\sqrt{h_{(1)}h_{(2)}} + h_{(2)}) \right] p_{h_{(1)}h_{(2)}}(h_{(1)}, h_{(2)}) dh_{(1)} dh_{(2)} dh_$$

Substituting Eq. (2.9) in Eq. (2.14), the expression of P_{avg}^2 can be simplified as

$$P_{\rm H}^{2} = \frac{C_{2}K(K-1)}{(h_{0})^{2}} \int_{h_{\rm th}^{2}}^{+\infty} \int_{h_{(2)}}^{+\infty} \left[\frac{2}{h_{(1)}} + 2\frac{1}{\sqrt{h_{(1)}h_{(2)}}} + \frac{1}{h_{(2)}} \right] \\ \times \exp\left(-\frac{h_{(1)}}{h_{0}}\right) \exp\left(-\frac{h_{(2)}}{h_{0}}\right) \left(1 - \exp\left(-\frac{h_{(2)}}{h_{0}}\right)\right)^{K-2} dh_{(1)} dh_{(2)} \\ = \frac{C_{2}K(K-1)}{(h_{0})^{2}} \sum_{i=0}^{K-2} {K-2 \choose i} (-1)^{i} \left[\int_{h_{\rm th}^{2}}^{+\infty} 2 \right] \\ \times \left[E_{1}\left(\frac{h_{(2)}}{h_{0}}\right) + E_{1/2}\left(\frac{h_{(2)}}{h_{0}}\right) \right] \exp\left(-\frac{(i+1)h_{(2)}}{h_{0}}\right) dh_{(2)} \\ + h_{0}E_{1}\left(\frac{(i+2)h_{\rm th}^{2}}{h_{0}}\right) \right]$$
(2.15)

where the function $E_n(\cdot)$ is defined as

$$E_n(x) = \int_1^{+\infty} \frac{\exp(-xt)}{t^n} dt.$$
 (2.16)

Similarly the average power in uniform 16-QAM constellation-based two user scheduling, P_u^2 can be written in simplified form as

$$P_{U}^{2} = \frac{C_{4}K(K-1)}{2(h_{0})^{2}} \sum_{i=0}^{K-2} {\binom{K-2}{i}} (-1)^{i} \left[\int_{h_{th}^{2}}^{+\infty} E_{1} \left(\frac{h_{(2)}}{h_{0}} \right) \right] \times \exp\left(-\frac{(i+1)h_{(2)}}{h_{0}} \right) dh_{(2)} + h_{0}E_{1} \left(\frac{(i+2)h_{th}^{2}}{h_{0}} \right) \right]$$
(2.17)

where the constant $C_4 = 2N_0\Theta(2^4 - 1)$. Unfortunately, the integrals in Eqs. (2.15) and (2.16) cannot be evaluated in simple closed-form. However, it can be evaluated numerically and consequently the average transmit power for TBS scheme can be obtained.

The average transmit power for single-user scheduling can be expressed as

$$\mathbf{P}^{1} = \int_{h_{\text{th}}^{1}}^{+\infty} \frac{C_{4}}{h_{(1)}} p_{h_{(1)}}(h_{(1)}) dh_{(1)}.$$
(2.18)



Figure 2.2: Normalized (by N_0) average transmit power versus the number of users.

Substituting Eq. (2.12) in Eq. (2.18), P¹ can be simplified to

$$P^{1} = \int_{h_{th}^{1}}^{+\infty} \frac{C_{4}}{h_{(1)}} \frac{K}{h_{0}} \exp\left(-\frac{h_{(1)}}{h_{0}}\right) \left(1 - \exp\left(-\frac{h_{(1)}}{h_{0}}\right)\right)^{K-1} dh_{(1)}$$

$$= \sum_{i=0}^{K-1} \frac{C_{4}K}{i+1} \binom{K-1}{i} (-1)^{i} E_{1}\left(\frac{(i+1)h_{th}^{1}}{h_{0}}\right).$$
(2.19)

The average power for different schemes is plotted in Fig. 2.2. There are two observations from this figure. First, the two user scheduling always requires more power than the single user scheduling. This is due to the inclusion of the second best user in the transmission. Second, the required additional power decreases as the number of users in the system increases. This can be explained by the fact that as the number of users increases in the system, the frequency of having higher difference between the first largest channel gain $h_{(1)}$ and the second largest channel gain $h_{(2)}$ decreases. Therefore, the average power required to transmit to the second best user decreases. On the other hand, the hierarchical schemes always require less power than the uniform constellation-based mini-slotted scheme. This is expected due to the superposition of the information bits for two different users in the same message symbol. It is also obvious from Fig. 2.2 that the required ad-

ditional power with mini-slotted scheme decreases as the number of users increases in the system. This is again expected, as the frequency of having higher difference between $h_{(1)}$ and $h_{(2)}$ decreases with the number of users in the system.

2.2.3 Buffer density and packet loss probability

In this section, we analyze and compare the buffer distribution, average buffer occupancy, and packet loss probability of all the schemes under consideration. We assume that each user has a transmission buffer where packets are queued until they are served to the desired users. The system is assumed to be a homogenous system where all users have the same (i) buffer length, (ii) packet arrival statistics, and (iii) average link gain. Therefore, we are interested in the queuing analysis of a given user's buffer. Other users' buffer will have the same performance. The finite buffer size is assumed to be *B* packets and the packet arrival process is assumed to be Bernoulli with arrival probability *a* in each time slot. We assume that a packet arriving during time slot n - 1 cannot be transmitted until time interval *n*.

Since with the two-best user scheduling schemes (both hierarchical and mini-slotted schemes) two packets are transmitted from one of the selected buffers per time slot, the number of packets transmitted during time interval n is min $(b_n, 2)$, where b_n is the buffer occupancy at time slot n. For i.i.d. channel variations among K users, the probability that a user will get an access to the channel, in given time slot, is 2/K. Therefore, the probability of serving a given queue, s_2 , for a pre-specified outage probability O_{out} , is expressed as³

$$s_2 = \frac{2}{K} (1 - O_{out}).$$
 (2.20)

The dynamics of the buffer with the proposed two-best user scheduling is shown in Fig. 2.3. The discrete time Markov chain describing the buffer dynamics has the following

³For the sake of simplicity of the analysis we assume the fading gain of a given user varies independently from one time slot to the next.



Figure 2.3: Markov chain representation of the buffer dynamic with power-controlled modulation-assisted two-best user scheduling.

transition matrix

where b = 1 - a and $v_2 = 1 - s_2$. The steady-state probability $\pi = [\pi_0 \ \pi_1 \ \cdots \ \pi_B]$ can obtained solving the following equations

$$\mathbf{b}_{\infty}\mathbf{P} = \boldsymbol{\pi} \tag{2.22}$$

$$\sum_{k=0}^{B} \pi_k = 1.$$
 (2.23)

The steady-state probability density function can be obtained using the standard procedures, such as the ones outlined in [35].

Similarly we can construct the Markov chain for the SBS scheme. Since four packets are taken out of the buffer per channel access, we assume that the number of packets trans-



Figure 2.4: Buffer distribution and average buffer occupancy level.

mitted during time interval n is min $(b_n, 4)$. In this case the probability s_1 that a given queue will be served in given time slot can be written as

$$s_1 = \frac{1}{K} (1 - O_{out}).$$
 (2.24)

The corresponding steady-state buffer distribution can be obtained using standard procedures. The steady-state buffer distribution function is plotted in Fig. 2.4. In the numerical example we consider K = 10, B = 10, and a = 0.1. From Fig. 2.4 we can observe that the buffer occupancy with power-controlled modulation-assisted TBS is concentrated in the low occupancy region with higher probability compared with the SBS. Therefore, powercontrolled TBS scheme results into a lower average buffer occupancy (consequently, lower average delay for the packets) which is also obvious from the marked points on the x-axis of Fig. 2.4. The buffer distribution which is concentrated in the low occupancy region with two best user scheduling also results into a lower packet loss probability due to the buffer overflow as we will see from the following analysis and example.

The packet loss probability, PLP defined as the probability that a packet is dropped due to buffer overflow corresponds to the probability that the buffer is full and a packet arrives



Figure 2.5: Packet loss probability versus packet arrival probability.

in the buffer. Therefore, the PLP can be written as

$$PLP = \pi_B a. \tag{2.25}$$

The packet loss probability with different schemes are shown in Fig. 2.5 for different values of packet arrival probability. This figure shows that the two best user scheduling always offer less packet loss probability. The is because of the comparatively lower probability of buffer occupancy in the high buffer occupancy region with two best user scheduling as mentioned above.

2.3 Conclusion

In this chapter, we proposed and studied a new hierarchical constellation-based multi-user opportunistic scheduling scheme. The multi-user scheduling provides lower average buffer occupancy (i.e., queuing delay) and loss probability for the packets at the transmission buffer at the expense of transmit power compared with the classical single user scheduling. The required additional power decreases as the number of users increases in the

system. The hierarchical scheme has been compared with a uniform constellation-based mini-slotted scheme. This mini-slotted scheme provides similar higher layer performances as the hierarchical scheme but requires higher transmit power. This additional power requirement decreases as the number of user increases in the system.

Bibliography

- [23] M. Morimoto, M. Okada, and S. Komaki, "A hierarchical image transmission system in fading channel," in *Proc. IEEE Int. Conf. Univ. Personal Commun. (ICUPC' 95)*, Tokyo, Japan, Oct. 1995, pp. 769-772.
- [24] P. K. Vitthaladevuni and M.-S. Alouini, "A recursive algorithm for the exact BER computation of generalized hierarchical QAM constellations," *IEEE Trans. Inform. Theory*, vol. 49, pp. 297-307, Jan. 2003.
- [25] M. B. Pursley and J. M. Shea, "Adaptive non-uniform phase-shift-key modulation for multimedia traffic in wireless networks," *IEEE Journal Select. Areas Commun.*, vol. 18, pp. 1394-1407, Aug. 2000.
- [26] Md. J. Hossain, P. K. Vitthaladevuni, M.-S. Alouini, and V. K. Bhargava, "Hierarchical modulations for multimedia and multicast transmission over fading channels," in *Proc. IEEE Veh. Technol. Conf. (VTC'03 Spring)*, Jeju, Korea, Apr. 2003, pp. 2633-2637.
- [27] T. Cover, "Broadcast channel," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 2-14, January 1972.
- [28] DVB-T standard: ETS 300 744., "Digital Broadcasting Systems for Television, Sound and Data Services: Framing Structure, Channel Coding and Modulation for Digital Terrestrial Television," (ETS Draft, 1.2.1, EN300 744, 1999-1).
- [29] FLO Forum and Qualcomm. Flo technology overview: Revolutionizing multimedia, http://www.floforum.org/, 2007. 3GPP2.
- [30] Ultra mobile broad physical layer, C.P0084-001, Feb. 2007.

- [31] S. Pietrzyk and G. J. M. Janssen, "Subcarrier and power allocation for QoS-aware OFDMA systems using embedded modulation," in *Proc. Int. Conf. Commun. (ICC'* 04), Paris, France, Jun. 2004, pp. 3202-3206.
- [32] A. J. Goldsmith, "The capacity of downlink fading channels using variable power and variable rate," *IEEE Trans. Veh. Technol.*, vol. 46, pp. 569-580, Aug. 1997.
- [33] J. M. Cioffi, "A multicarrier premier," November 1991, also available in http://www.stanford.edu/group/cioffi/pdf/multicarrier.pdf
- [34] T. Eng, N. Kong, and L. B. Milstein, "Selection combining schemes for RAKE receivers," in *Proc. 4th Int. Conf. Univ. Personal Commun. (ICUPC'95)*, Tokyo, Japan, Nov. 1995, pp. 426-430.
- [35] F. Gebali, Computer Communication Networks: Analysis and Design, NorthStar Digital Design, Inc., Victoria, BC, Canada, 2002.

Chapter 3

Rate Adaptive Hierarchical Modulation-Assisted Two-User Opportunistic Scheduling

⁴Inspired by the power efficient hierarchical multiplexing schemes, in previous chapter (also refer to [36]), we studied a *multi-user* opportunistic scheduling scheme using a variable power fixed size hierarchical constellation. However, due to the inclusion of second best user in each transmission, the scheme proposed in [36] offered more frequent information access to the users at the expense of certain amount of transmit power compared with the classical SBS scheme. In this chapter, we propose and study a two-best user opportunistic scheduling scheme using fixed power discrete rate adaptive hierarchical constellations. Rather than including the second best user in each transmission as in [36], the newly proposed scheme judiciously decides whether the second best can be included in a given transmission or not. As such the frequency of channel access of the users increases without any expense of overall average SE compared with the classical SBS scheme. Specifically, according to the channel quality of the first best user, the alphabet size of the square quadrature amplitude modulation (QAM) is selected in order to maximize the SE. However, if the first best user's channel quality is good enough to decode the information transmitted in the second hierarchy of the selected size hierarchical QAM constellation while the second best user's channel quality is capable of decoding the information transmitted in the first hierarchy, information is transmitted to both the selected first and second best users using the hierarchical QAM constellation. Otherwise, information is transmitted only to the first best user using the selected size uniform QAM constellation. We present expressions for the

⁴The research work presented in this chapter has been published as Md. J. Hossain, M.-S. Alouini, and V. K. Bhargava, "Rate Adaptive Hierarchical Modulation-Assisted Two-User Opportunistic Scheduling", *IEEE Trans. Wireless Commun.*, Vol. pp. 2076-2085, Jun. 2007.

information transmission probability and the overall average SE [Bits/Sec/Hz] defined as the overall average transmitted data rate per unit bandwidth for a specified number of users, a specified average CNR, and BER. Numerical results for a fading environment where the channel variation among the users is i.i.d. are presented. The novelty of the fixed power discrete rate adaptive hierarchical modulation-assisted two-best user scheduling is that it provides more frequent access to the information (almost double) without any average SE degradation compared with the classical single best user opportunistic scheduling.

3.1 Channel Model and Hierarchical Constellations

3.1.1 Channel model

Let the total system bandwidth be W[Hz]. Given the channel fading amplitude of kth user, α_k and a white noise power spectral density of $N_0[W/\text{Hz}]$, let us define the normalized fading gain at receiver for kth user $h_k = (\alpha_k)^2/N_0W$. For a transmit power of P[W], the received symbol CNR of kth user's receiver $g_k = h_k P$. We consider that the fading amplitude of the users is characterized by the Rayleigh PDF which leads to the following PDF of the kth user received symbol CNR g_k , $p_{g_k}(g_k)$

$$p_{g_k}(g_k) = \frac{1}{\bar{g}_k} \exp\left(-\frac{g_k}{\bar{g}_k}\right), \quad g_k \ge 0$$
(3.1)

where \bar{g}_k is the average CNR of user k.

For total K users whose channels vary identically and independently, the PDF of the largest CNR $g_{(1)}$, $p_{g_{(1)}}(g_{(1)})$ can be expressed as [39]

$$p_{g_{(1)}}(g_{(1)}) = \frac{K}{g_0} \exp\left(-\frac{g_{(1)}}{g_0}\right) \left(1 - \exp\left(-\frac{g_{(1)}}{g_0}\right)\right)^{K-1}, \quad g_{(1)} \ge 0$$
(3.2)

where g_0 is the identical average CNR of the users. The joint PDF of the two largest CNRs $g_{(1)}$ and $g_{(2)}$, $p_{g_{(1)}g_{(2)}}(g_{(1)}, g_{(2)})$ can be expressed as [39]

$$p_{g_{(1)}g_{(2)}}(g_{(1)},g_{(2)}) = \frac{K(K-1)}{(g_0)^2} \exp\left(-\frac{g_{(1)}}{g_0}\right) \exp\left(-\frac{g_{(2)}}{g_0}\right) \left(1 - \exp\left(-\frac{g_{(2)}}{g_0}\right)\right)^{K-2}$$
(3.3)

where $g_{(1)} \ge g_{(2)} \ge 0$.



Figure 3.1: Hierarchical 4/64-QAM constellation.

3.1.2 Hierarchical constellations

Due to their inherent SE square QAM constellations have received considerable attention in mobile communication (see, e.g., [38]). We consider only square $M = 2^{2r}$ -QAM constellations, where $r = 1, 2, \dots, Q$, in our proposed scheme.

The detailed description of *generalized* hierarchical constellations can be found, for example, in [37]. Briefly a generalized hierarchical $4/16/\cdots/M(=2^{2r})$ QAM constellation has m levels of hierarchies and the relative distances between the symbols within the constellation are described by a set of r distances [37]. The distance set can be represented as a vector [37]

$$\vec{d} = [d_1 \ d_2 \cdots d_r] \tag{3.4}$$

where d_k is the distances between the symbols in kth hierarchy. Since our proposed scheme transmits information to two best users using hierarchical constellation, we use a family of hierarchical $N(=2^{2s})/M(=2^{2r})$ -QAM constellations, where $r = 2, 3, \dots, Q$ and $s = 0, 1, \dots, r - 1$. We assume that there are two incoming streams of data where one stream has higher priority than the other. With hierarchical $N(=2^{2s})/M(=2^{2r})$ -QAM constellation, the first priority 2s bits are assigned to the most significant s bit positions in the I and Q phase channels. The lowest priority 2(r - s) information bits are assigned to the rest of the positions in the in-phase and quadrature-phase. For example, see 4/64-QAM constellation with Karnaugh map style Gray mapping in Fig. 3.1. In this figure, one of the fictitious 4-QAM symbols (denoted by the grey circle) is selected by the highest priority two bits. The black symbols are the actual transmitted symbols. The one of the black 16-QAM symbols surrounding the selected 4-QAM symbol is choosen by the least priority four bits.

In order to avoid confusion between the generalized $4/16 \cdots /2^{2r}$ -QAM and $2^{2s}/2^{2r}$ -QAM constellations, it is important to mention that the first s hierarchies of the generalized $4/16/\cdots/2^{2r}$ -QAM constellation are considered as one hierarchy in $2^{2s}/2^{2r}$ -QAM constellation which we refer as the first hierarchy (namely hierarchy H₁). The distances between these s (1 < s < r) hierarchies are uniformly spaced i.e., [37]

$$d_{i-1} = 2d_i, \quad i = s, s - 1, \dots, 2.$$
 (3.5)

The rest r - s hierarchies of the generalized hierarchical constellation are considered as another hierarchy which is referred as the second hierarchy (namely hierarchy H₂). The distances between these r - s hierarchies are also uniformly spaced i.e.,

$$d_{i+1} = \frac{d_i}{2}, \quad i = s+1, s+2, \cdots, r-1.$$
 (3.6)

Therefore the relative symbol positions within the hierarchical $2^{2s}/2^{2r}$ -QAM constellation are defined by two distances; d_s and d_{s+1} in contrast to r distances in generalized hierarchical $4/16/\cdots/2^{2r}$ -QAM constellation. A constellation priority parameter p can be defined as

$$p = \frac{d_s}{d_{s+1}}.\tag{3.7}$$

For uniform square QAM constellations, the value of p is two (see, e.g., [37]). For a given constellation size $M(=2^{2r})$, there can be a set of hierarchical constellations with different number of bits in each of the two hierarchies as described above. As such the set is denoted by $C_r = \{2^2/2^{2r}, 2^4/2^{2r} \cdots, 2^{2(r-1)}/2^{2r}\}$. Later on we will use the set of the hierarchical constellations of a given size to allocate the available transmission rate among the selected two users.

The exact BER expressions over AWGN channels were shown in [37] as the summation of weighted Q-functions and to be solely dependent on the constellation size M, the CNR g, and the priority parameter p. Therefore, in general, the BER expressions for the bits in the first hierarchy, BER₁ can be written as

$$BER_1 = f_1(M, p, g).$$
(3.8)

Similarly, the BER expressions for the bits in the second hierarchy, BER₂ can be written as

$$BER_2 = f_2(M, p, g).$$
 (3.9)

3.2 Two Best-User Opportunistic Scheduling

3.2.1 General assumptions

Our proposed scheme has the following operating assumptions:

A1: A single BS is transmitting information to K users opportunistically.

A2: The channel variation among the users is independent.

A3: The channel qualities of the users are reported to the BS in the form of instantaneous CNR via the feedback channel.

A4: The users have a target PER, PER₀ and a FEC strategy is used. Specifically, the packets are encoded with block coding for example, BCH code which is capable of correcting up to t errors. For a given FEC scheme with capability of correcting t errors, the target BER, BER₀ in order to meet the target PER, PER₀ can be obtained easily if the error event is random. This random error event can be achieved by using proper interleaving.

3.2.2 Overall system description

The block diagram of our proposed rate adaptive hierarchical modulation-assisted TBS is shown in Fig. 3.2. In what follows, we describe the general working principle of our proposed scheme. According to the available instantaneous CNR estimate of all the users, the BS scheduler selects two users, namely the *first best user* (denoted as $U_{(1)}$) and the *second best user* (denoted as $U_{(2)}$), whose instantaneous CNRs are the first and second highest, respectively. Without loss of generality let us denote, the CNRs of user $U_{(1)}$ and



Figure 3.2: System model of BS scheduler.

 $U_{(2)}$ by $g_{(1)}$ and $g_{(2)}$, respectively with $g_{(1)} \ge g_{(2)}$. According to the first best user's channel CNR, $g_{(1)}$, the size of the constant power QAM constellation is selected by the constellation size selector as shown in Fig. 3.2. Next, jointly considering the channel CNRs of both the selected best users, $U_{(1)}$ and $U_{(2)}$, the bit assigner (see Fig. 3.2) decides how the available total transmission rate is divided among the selected users $U_{(1)}$ and $U_{(2)}$. The constellation size as well as the bit assignment procedure will be described in details in the next section.

When information bits are transmitted to both the users $U_{(1)}$ and $U_{(2)}$, a hierarchical N/M-QAM is used and since first hierarchy is more protected than the second hierarchy, the second best user's transmission is scheduled in the first hierarchy whereas the first best user's transmission is scheduled in the second one. The decision i.e., the position and the number of bits in the transmitted symbol for the selected users, is informed to the users' receiver via the control channels. Therefore, the selected users receive the symbols and look only for the bits in particular positions within the symbol.

3.2.3 Rate adaptation process

In this section, we describe how the constellation size is selected according to the first best user channel quality and how the total available transmission rate is allocated between the two best users jointly considering their instantaneous link qualities. For convenience the rate adaptation process of our rate adaptive TBS scheme is summarized via a flowchart in Fig. 3.3.

Discrete constellation size selection

To minimize the power consumption, the proposed scheme does not attempt to transmit when the first best user's instantaneous CNR, $g_{(1)}$ is not good enough to support the satisfactory communication using the 4-QAM constellation. In other words, the transmission is suspended if the estimated CNR, $g_{(1)}$ is less than γ_1 (as shown in Fig. 3.3) where γ_1 is obtained from the exact BER of uniform 4-QAM modulation in AWGN⁵ given by

$$BER_{4-QAM} = \frac{1}{2} erfc(\sqrt{g_{(1)}}).$$
 (3.10)

In order to maximize the overall SE our proposed scheme selects the constellation size according to the first best user's received CNR as follows. The CNR range of the user $U_{(1)}$, $g_{(1)}$ is divided into Q + 1 fading regions; namely $r = 0, 1, 2, \dots, Q$. Each fading region is assigned a unique constellation size. Specifically, the constellation size $M_r = 2^{2r}$ (where 2r is the total number of bits per M_r -QAM symbol) is assigned to the *r*th fading region $(r = 0, 1, 2, \dots, Q)$. When the received CNR of the user $U_{(1)}$, $g_{(1)}$ is estimated to be in the *r*th fading region, the constellation size M_r is selected to be transmitted as shown in Fig. 3.3. Choosing the constellation size M_0 in region r = 0 corresponds to no data transmission. The region boundaries (or constellation size switching thresholds) $\{\gamma_r\}_{r=1}^{Q+1}$ are set to the CNR such that the information bits are sent below the target BER, BER₀ using the uniform M_r -QAM constellation. This is very similar to a conventional discrete rate adaptive transmission system [38]. Since the exact BER expressions for uncoded M-QAM in AWGN [37] are not invertible, the region boundaries γ_r ($r = 1, 2, \dots, Q + 1$) are obtained by numerically solving the BER equations of the uniform M_r -QAM in AWGN for CNR and setting $\gamma_{Q+1} = +\infty$.

⁵In all the schemes under consideration, the choices of constellation size and bit assignments to the users are done in a conservative fashion. Specifically, the alphabet size M is selected and the available transmission rate is divided among the two best users such that their instantaneous channel qualities support the bits below the target BER. In addition, we assume that the fading gains remain constant over the symbol period. Therefore, the BER expressions in AWGN are used to set the switching thresholds.



Figure 3.3: Flow chart of proposed rate adaptive modulation-assisted TBS scheme.

Transmission rate allocation to the selected users $U_{(1)}$ and $U_{(2)}$

To transmit information to the selected second best user opportunistically, the scheme switches between the uniform and hierarchical constellations jointly considering the CNR estimate fluctuation of the two best users as follows. As we have described in Section 3.1.2 that a QAM constellation of order $M_r = 2^{2r}$, $(r = 2, 3, \dots, Q)$ can form a set of hierarchical constellations C_r . Each hierarchical constellation in the set, C_r specifies a unique combinations of bits in two different hierarchies. Therefore, in a given fading region, the number of bits within the constellation that can be assigned to the users $U_{(1)}$ and $U_{(2)}$ jointly depends on the received CNR estimate of both the users. To assign the available transmission rate (i.e., the number of the bits) to the selected users, $U_{(1)}$ and $U_{(2)}$, each of the fading regions, except r = 1, is divided into a number of fading sub-regions and each sub-region supports a unique combination of bits in the first and second hierarchies. Specifically, the rth fading region $(r = 2, 3, \dots, Q)$ is divided into $s = 0, 1, \dots, r-1$ sub-regions and when the CNR of the first best user is estimated in the sth sub-region s, 2(r-s) bits in the second hierarchy can be decoded by the first best users with required reliability⁶. The subregion boundaries $\{\gamma_{r,s}^2\}_{r=2,s=1}^{Q,r-1}$ are obtained by numerically solving the BER equations of the hierarchical $2^{2s}/2^{2r}$ -QAM such that the bits assigned in the second hierarchy are sent below the target BER, BER₀. For example, the sub-region boundary $\gamma_{3,2}^2$ for the 2nd subregion of the 3rd fading region are obtained by numerically solving (3.9) for the CNR such that

$$f_2(M_3, p, \gamma_{3,2}^2) = \text{BER}_0.$$
 (3.11)

Although the first best user can decode the 2(r - s) information bits transmitted in the second hierarchy of the hierarchical $2^{2s}/2^{2r}$ -QAM with given reliability if its CNR estimate falls in the *s*th subregion of *r*th fading region, the 2*s* bits transmitted in the first hierarchy may not be decoded by the second best user with the required reliability. It depends on the CNR of the second best user, $g_{(2)}$. Therefore, in order to decide whether to transmit information to the user $U_{(2)}$ or not, the bit assigners checks the channel CNR, $g_{(2)}$. If $g_{(2)}$ is good enough to decode the 2*s* information bits transmitted in the first hierarchy with the required reliability, the bit assigner assigns these bits to the user $U_{(2)}$. Consequently, hierar-

⁶In sub-region 0 of a given region r corresponds to transmission to only the first best user using uniform 2^{2r} -QAM.

chical $2^{2s}/2^{2r}$ -QAM is transmitted from the demodulator as shown in Fig. 3.3. Otherwise, the available total transmission rate r is assigned to the user $U_{(1)}$ using the uniform 2^{2r} -QAM. Therefore, if the channel CNR, $g_{(1)}$ falls in a given sub-region, the demodulator can transmit either the uniform constellation or the hierarchical constellation depending on the CNR estimate $g_{(2)}$. The switching between the uniform and the hierarchical constellations in a sub-region s ($s = 1, 2, \dots, r - 1$) of the fading region r ($r = 2, 3 \dots, Q$) depends on the CNR, $g_{(2)}$ and is determined by the thresholds, $\{\gamma_{r,s}^1\}_{r=2,s=1}^{Q,r-1}$. These thresholds $\gamma_{r,s}^1$ are selected such that the s bits assigned in the first hierarchy of $2^{2s}/2^{2r}$ -QAM are transmitted below the target BER, BER₀. For example, the switching threshold $\gamma_{3,2}^1$ in the second subregion of the 3rd fading region are obtained by numerically solving Eq. (3.8) for the CNR such that

$$f_1(M_3, p, \gamma_{3,2}^1) = \text{BER}_0. \tag{3.12}$$

For instance, the details of bit assignment to the selected users in different sub-regions for a maximum constellation size of 8 bits (i.e., 256-QAM) are summarized in Fig. 3.4. The joint PDF of the channel CNRs, $g_{(1)}$ and $g_{(2)}$ is a two dimensional space, the valid region is defined by the condition: $g_{(1)} \ge g_{(2)} \ge 0$ as shown in Fig. 3.4. This valid region is divided into a number of boxes. The region boundaries of each box are determined by the thresholds which are shown symbolically. Each box is assigned with a constellation (either uniform or hierarchical) as written on the box. In the region where $2^{2s}/2^{2r}$ -QAM constellation is transmitted information is transmitted to both users $U_{(1)}$ and $U_{(2)}$ while in the region where 2^{2r} -QAM constellation is transmitted, information is transmitted only to user U₍₁₎. For example, if the channel CNR, $g_{(1)}$ is above $\gamma_{4,2}^2$ but less than $\gamma_{4,3}^2$, information is transmitted both the selected users using hierarchical 16/256-QAM provided that the channel CNR $g_{(2)}$ falls above $\gamma_{4,2}^1$. If $g_{(2)}$ is less than $\gamma_{4,2}^1$ although $g_{(1)}$ falls in the above mentioned sub-region, no data is transmitted to the second best user because it cannot decode the information bits transmitted in the first hierarchy. In this case, information is only transmitted to the first best user using the uniform 256-QAM constellation as shown in Fig. 3.4. The figure shows that in the 0th sub-region of any fading region, no data is transmitted to the second best user because the first best user's channel quality is not good enough to decode the information transmitted in the second hierarchy. In fading region (r = 1), the scheme proposes to transmit 4-QAM which has only one hierarchy. However, it has two different phases; I and Q phases. If the second best user's CNR can support 4-



Figure 3.4: Example: Joint CNRs to constellation mapping.

QAM as well, an information bit is transmitted in one of the phases for the first best user and an information bit is transmitted in the other phase for the second best user.

3.3 Performance Analysis

In this section, we derive analytical expressions for the average spectral efficiencies and the information transmission probability. Numerical results obtained via computer simulation are also presented for an i.i.d Rayleigh fading environment. A target BER, BER₀ of 10^{-4} is used. We use the MATLAB Communication Toolbox to obtain the numerical results. We assume perfect channel estimation, coherent phase detection at the receiver, and Gray coding for bit mapping on the signal constellations, as shown in Fig. 3.1. In the examples, we use a maximum constellation size of 1024 for both hierarchical and uniform QAM constellations.

3.3.1 Average spectral efficiency

The probabilities, $a_{r,s}$ that the $2^{2s}/2^{2r}$ -QAM constellations are transmitted i.e., the CNR $g_{(2)}$ is above the threshold $\gamma_{r,s}^1$ while the CNR $g_{(1)}$ falls in sth sub-region in the rth (r > 1) fading region is given by

$$a_{r,s} = \begin{cases} \int_{\gamma_{r,s}^{2}}^{\gamma_{r,s+1}^{2}} \int_{\gamma_{r,s}^{1}}^{g_{(1)}} p_{g_{(1)}g_{(2)}}(g_{(1)}, g_{(2)}) dg_{(2)} dg_{(1)} & s \neq r-1 \\ \\ \\ \int_{\gamma_{r,s}^{2}}^{\gamma_{r+1}^{2}} \int_{\gamma_{r,s}^{1}}^{g_{(1)}} p_{g_{(1)}g_{(2)}}(g_{(1)}, g_{(2)}) dg_{(2)} dg_{(1)} & s = r-1. \end{cases}$$
(3.13)

Using the expressions of the joint PDF, $p_{g_{(1)}g_{(2)}}(g_{(1)}, g_{(2)})$ of the largest two CNRs for i.i.d. case in Eq. (3.3), the integrals in Eq. (3.13) can be evaluated in closed-form as shown in Appendix A. Using the results in Appendix A and after some manipulations, the probabilities $\{a_{r,s}\}_{s=1}^{s=r-2}$ in Eq. (3.13) can be expressed in closed-form as follows:

$$a_{r,s} = \sum_{k=0}^{K-2} \frac{K(K-1)}{k+1} \binom{K-2}{k} (-1)^k \left[A(\gamma_{r,s+1}^2) \exp\left(-\frac{(k+1)\gamma_{r,s}^1}{g_0}\right) - \frac{B(\gamma_{r,s+1}^2)}{k+2} \right]$$
(3.14)

where

$$A(x) = \exp\left(-\frac{\gamma_{r,s}^2}{g_0}\right) - \exp\left(-\frac{x}{g_0}\right), \qquad (3.15)$$

and

$$B(x) = \exp\left(-\frac{(k+2)\gamma_{r,s}^2}{g_0}\right) - \exp\left(-\frac{(k+2)x}{g_0}\right).$$
 (3.16)

Similarly, the probabilities $\{a_{r,s}\}_{s=r-1}$ in Eq. (3.13) can be expressed in closed-form as follows:

$$a_{r,s} = \sum_{k=0}^{K-2} \frac{K(K-1)}{k+1} \binom{K-2}{k} (-1)^k \left[A(\gamma_{r+1}^2) \exp\left(-\frac{(k+1)\gamma_{r,s}^1}{g_0}\right) - \frac{B(\gamma_{r+1}^2)}{k+2} \right].$$
(3.17)

Recall from Section 3.2.3 that in a given region r ($r = 2, 3, \dots, Q$), in two situations, information is transmitted only to the first best user U₍₁₎ using uniform 2^{2r} -QAM; if the CNR $g_{(1)}$ falls in the 0th sub-region or the CNR $g_{(2)}$ is less than the threshold $\gamma_{r,s}^1$ although the CNR $g_{(1)}$ falls in sth sub-region in the rth fading region. Therefore, the total probability, $a_{r,0}$ that the uniform 2^{2r} -QAM constellation is transmitted in the rth fading region is the summation of all the individual probabilities and can be written as

$$a_{r,0} = \int_{\gamma_r}^{\gamma_{r,1}^2} p_{g_{(1)}}(g_{(1)}) dg_{(1)} + \sum_{s=1}^{r-1} \int_{\gamma_{r,s}^2}^{\gamma_{r,s+1}^2} \int_0^{\gamma_{r,s}^1} p_{g_{(1)}g_{(2)}}(g_{(1)}, g_{(2)}) dg_{(2)} dg_{(1)}.$$
 (3.18)

Again using the expressions of the PDF $p_{g_{(1)}}(g_{(1)})$ and the joint PDF, $p_{g_{(1)}g_{(2)}}(g_{(1)}, g_{(2)})$ for i.i.d. case in Eqs. (3.3) and (3.2), respectively, Eq. (3.18) can be expressed in closed-form as (see Appendix 9.1.5 for details):

$$a_{r,0} = \sum_{k=0}^{K-1} \frac{K}{k+1} \binom{K-1}{k} (-1)^k \left[\exp\left(-\frac{(k+1)\gamma_r}{g_0}\right) - \exp\left(-\frac{(k+1)\gamma_{r,1}^2}{g_0}\right) \right] \\ + \sum_{s=1}^{r-1} \left[\left[\exp\left(-\frac{\gamma_{r,s}^2}{g_0}\right) - \exp\left(-\frac{\gamma_{r,s+1}^2}{g_0}\right) \right] \\ \times \left[\sum_{k=0}^{K-2} \frac{K(K-1)}{k+1} \binom{K-2}{k} (-1)^k \left[1 - \exp\left(-\frac{(k+1)\gamma_{r,s}^1}{g_0}\right) \right] \right].$$
(3.19)

The probability of transmitting information only to the user $U_{(1)}$ using uniform 4-QAM constellation corresponds to the probability that the CNR $g_{(1)}$ falls in the region, r = 1 but not the CNR $g_{(2)}$ and is given by

$$a_{1,0} = \int_{\gamma_1}^{\gamma_2} \int_0^{\gamma_1} p_{g_{(1)}g_{(2)}}(g_{(1)}, g_{(2)}) dg_{(2)} dg_{(1)}.$$
(3.20)

Using Eq. (3.3), the probability in Eq. (3.20) for i.i.d case can be expressed as

$$a_{1,0} = \left[\exp\left(-\frac{\gamma_1}{g_0}\right) - \exp\left(-\frac{\gamma_2}{g_0}\right) \right] \sum_{k=0}^{K-2} \frac{K(K-1)}{k+1} \binom{K-2}{k} (-1)^k \\ \times \left[1 - \exp\left(-\frac{(k+1)\gamma_1}{g_0}\right) \right].$$
(3.21)

Similarly, the probability $a_{1,1}$ that information is transmitted to both the selected users using uniform 4-QAM constellation can be expressed as

$$a_{1,1} = \int_{\gamma_1}^{\gamma_2} \int_{\gamma_1}^{g_{(1)}} p_{g_{(1)}g_{(2)}}(g_{(1)}, g_{(2)}) dg_{(2)} dg_{(1)}.$$
(3.22)

For the i.i.d. case, the integral in Eq. (3.22) can be evaluated in closed-form (see the Appendix 9.1.5).

$$a_{1,1} = \sum_{k=0}^{K-2} \frac{K(K-1)}{k+1} {\binom{K-2}{k}} (-1)^k \frac{k+1}{k+2} \exp\left(-\frac{(k+2)\gamma_1}{g_0}\right) \\ -\exp\left(-\frac{\gamma_2}{g_0}\right) \left[\exp\left(-\frac{(k+1)\gamma_1}{g_0}\right) + \frac{\exp\left(-\frac{(k+1)\gamma_2}{g_0}\right)}{k+2}\right].$$
(3.23)

The average link spectral efficiencies $\langle R_{(i)} \rangle / W$ (i = 1, 2) for information transmission to the users U₍₁₎ and U₍₂₎, respectively, are just the sum of data rates associated with the individual sub-regions of each region, weighted by the probabilities $a_{r,s}$ and can be expressed as

$$\frac{\langle R_{(1)} \rangle}{W} = \sum_{r=1}^{Q} \sum_{s=0}^{r-1} 2(r-s) a_{r,s} + a_{1,1}, \qquad (3.24)$$

$$\frac{\langle R_{(2)} \rangle}{W} = \sum_{r=2}^{Q} \sum_{s=1}^{r-1} 2s \, a_{r,s} + a_{1,1}.$$
(3.25)

Using Eqs. (3.14), (3.17), (3.21) and (3.23), the average spectral efficiencies in Eqs. (3.24) and (3.25) for the i.i.d. case can be expressed in closed-form. For the two-best user opportunistic scheduling the sum average SE can be obtained by adding the individual SE of the selected users. The sum average SE $\langle R^2 \rangle / W$ is, therefore, given by

$$\frac{\langle R^2 \rangle}{W} = \frac{\langle R_{(1)} \rangle}{W} + \frac{\langle R_{(2)} \rangle}{W}.$$
(3.26)

In SBS the overall average SE using discrete rate adaptive square uniform QAM constellations, $\langle R^1 \rangle / W$ is given by

$$\frac{\langle R^1 \rangle}{W} = \sum_{r=0}^{Q} 2r \ a_r \tag{3.27}$$

where a_r are the probabilities that the CNR $g_{(1)}$ falls in the regions r $(r = 0, 1, \dots, Q)$ and can be expressed as

$$a_r = \int_{\gamma_r}^{\gamma_{r+1}} p_{g_{(1)}}(g_{(1)}) dg_{(1)}.$$
(3.28)

Using Eq. (3.2), Eq. (3.28) can be expressed in desired closed-form as

$$a_{r} = \sum_{k=0}^{K-1} \frac{K}{k+1} \binom{K-1}{k} (-1)^{k} \left(\exp\left(-(k+1)\gamma_{r}/g_{0}\right) - \exp\left(-(k+1)\gamma_{r+1}/g_{0}\right) \right).$$
(3.29)

For a given constellation parameter, p the achievable overall average SE with our proposed rate adaptive TBS scheme, in an i.i.d. fading environment, is plotted in Fig. 3.5. The achievable SE in SBS is also plotted in this figure. According to classical theory, one would expect a degradation of the overall average SE if second best user is included in a transmission. However, from Fig. 3.5, we observe that there is no degradation of overall average SE with our proposed two-best user scheduling scheme. This is because our proposed scheme selects the constellation size according to the first best user's channel CNR in order to maximize the SE while satisfying a target BER. Then second best user is included in the transmission if and only if its channel quality is good enough to support the information transmitted in the first hierarchy of the selected size hierarchical constellation while the first best user can decode information transmitted in the second hierarchy with required reliability. Otherwise information is transmitted only to the first best user using the selected size uniform QAM constellation. In other words, our proposed scheme exploits the finite CNR gap between the two rate switching thresholds in conjunction with hierarchical constellations to include the second best user in a transmission. Therefore, inclusion of the second best user in a transmission system does not affect the selected discrete constellation size (i.e., the instantaneous SE) at all and two-best user scheduling experiences no degradation of overall average SE compared with the single best user scheduling. It is also obvious from Fig. 3.5 that the overall SE increases as the number of users increases in the system for a given value of the average CNR, g_0 . This is due to the multiuser diversity gain. The overall SE also increases as the average CNR of the users increase for a given number of users in the system, as expected from the concept of adaptive modulation (e.g., see [38]).

3.3.2 Access probability

Since the proposed scheme transmits information to the first best user irrespective of the second best user if its channel CNR $g_{(1)}$ is above the threshold γ_1 , the probability, $P_{a:1}$ that the selected first best user gets information access *with or without second best user* is given by

$$\mathbf{P}_{a:1} = \int_{\gamma_1}^{+\infty} p_{g_{(1)}}(g_{(1)}) dg_{(1)}.$$
(3.30)

Using Eq. (3.2), the probability, $P_{a:1}$ in Eq. (3.30) can be written in desired closed-form as

$$\mathbf{P}_{a:1} = \sum_{k=0}^{K-1} {\binom{K-1}{k}} (-1)^k \frac{K}{k+1} \exp\left(-(k+1)\gamma_1/g_0\right).$$
(3.31)

For i.i.d. channel variations among K users, the probability that a user will be selected to be transmitted in single best user scheduling is 1/K. Therefore, if discrete rate adaptive QAM constellations are used in single user scheduling, the information access probability of a given user is expressed as

$$P_{\text{access}}^{1} = \frac{1}{K} \sum_{k=0}^{K-1} {\binom{K-1}{k}} (-1)^{k} \frac{K}{k+1} \exp\left(-(k+1)\gamma_{1}/g_{0}\right).$$
(3.32)

The probability of transmitting information *simultaneously* to both selected users $U_{(1)}$ and $U_{(2)}$, is the sum of all the individual information transmission probabilities to these users in each fading sub-region. These probabilities are denoted by $\{a_{r,s}\}_{r=1,s=1}^{Q,r-1}$ and consequently the information access probability of the users $U_{(1)}$ and $U_{(2)}$ simultaneously, P_{s_2} can be expressed as

$$\mathbf{P}_{s_2} = \sum_{r=1}^{Q} \sum_{s=1}^{r-1} a_{r,s}.$$
(3.33)

Using the values of $\{a_{r,s}\}_{r=1,s=1}^{Q,r-1}$ which have been derived in Section IV-A, the probability P_{s_2} in Eq. (3.33) can be evaluated in desired closed-form.

The probability of transmitting information to the selected first best user, $U_{(1)}$ without user $U_{(2)}$ is equal to the probability, $P_{a:1}$ that the selected first best user gets information access with and without user $U_{(2)}$ minus the probability, P_{s_2} that both users $U_{(1)}$ and $U_{(2)}$ get information access simultaneously. This probability denoted by P_{S_1} can be written as

$$P_{s_1} = P_{a:1} - P_{s_2}. \tag{3.34}$$



Figure 3.5: Overall SE versus number of users (SE curves for single and two-best user scheduling overlap).



Figure 3.6: Information access probability versus number of users.

For i.i.d. channel variations among K users, the probability that a user will be selected to be transmitted in two-best user scheduling is 2/K. Therefore, the probability of information access by a user, P_{access}^2 is expressed as

$$\mathbf{P}_{\text{access}}^2 = \frac{1}{K} \mathbf{P}_{s_1} + \frac{2}{K} \mathbf{P}_{s_2}.$$
 (3.35)

Using Eqs. (3.32), (3.33) and (3.34) and after some manipulations, Eq. (3.35) can be expressed as

$$P_{access}^{2} = \frac{1}{K} \left[\sum_{k=0}^{K-1} {\binom{K-1}{k}} (-1)^{k} \frac{K}{k+1} \exp\left(-(k+1)\gamma_{1}/g_{0}\right) + \sum_{r=1}^{Q} \sum_{s=1}^{r-1} a_{r,s} \right].$$
 (3.36)

The information transmission probabilities with single and two-best user opportunistic scheduling schemes are plotted in Fig. 3.6. This figure shows that in an i.i.d. fading environment our proposed two-user scheduling offers almost double information access probability compared with the conventional single best user scheduling. This result is expected because the proposed scheme includes the second best users in the transmission. The figure also shows that the information access probability changes as the average CNR of the users is varied. Since the proposed scheme schedules opportunistically the second best user, by jointly considering the first and second best users' channel CNRs (as shown in Fig. 3.4), the access probability depends on the shape of the joint PDF. Obviously, the shape changes with different values of the average CNR and consequently, the access probability changes. In addition, as expected, the access probability also decreases with the number of users.

3.3.3 Effect of constellation parameter

As we have described in Section 3.1.2 that each hierarchical constellation has a priority parameter, p which controls the relative positions of the symbols within the constellation. The relative importance level of the bits transmitted in different hierarchies can be varied by changing the parameter p as illustrated in [37]. Equivalently, the two hierarchies require two different CNRs in order to meet a given target BER. These CNR thresholds can also be varied by changing the parameter p. In this section, we study the effect of the priority parameter on the average SE and the information access probability. Since the alphabet size of uniform QAM constellations is selected according to the first best user's CNR, there is



Figure 3.7: Average SE versus priority parameter p ($g_0 = 15$ [dB]).

no effect of p on the overall average SE. However, the parameter p has an effect on the individual SE of the first and second best which are plotted in Fig. 3.7. This figure shows that for a given number of users in the system as the value of p increases, the second best user's average SE decreases while that of first best user increases. We can also observe that the SE curves of the first best user with different number of users' information access probability as well as the SE depends on the values of switching thresholds, $\gamma_{r,s}^{1/2}$ and the shape of the two dimensional joint PDF $p_{g_{(1)}g_{(2)}}(g_{(1)}, g_{(2)})$ with respect to the switching thresholds. The shape of the joint PDF changes as the number of users changes in the system. On the other hand, if the values of constellation parameter p are changed, the values of the thresholds are changed. Therefore, for some values of K and p, it can happen that the location and shape of the joint PDF in $g_{(1)} - g_{(2)}$ plane and the switching thresholds are in a good fit which results in higher transmission opportunity for the first best user than that with some other values of K and p. As such the SE for the first best user intersects with each other for different values of K and p as observed in Fig. 3.7.

The information probability of the first best user does not change with p as information



Figure 3.8: Second best user information access probability versus priority parameter p ($g_0 = 15$ [dB]).

is transmitted to the first best user as soon as it's channel quality is good enough to support 4-QAM. However, the information access probability of the second best user is changed with p as shown in Fig. 3.8. Since the first best user's information access probability does not depend on p, maximizing the second best user's information access probability with respect to p maximizes the overall information access probability of a given user. Fig. 3.8 shows that there is a specific value of p for which second best user has a maximum value of information access probability and consequently the overall information access of a given user is maximized. There is also intersection between the curves of information access probability of user U₍₂₎ for different values of K as observed in Fig. 3.8. The main reason of this intersection is similar as the reason explained for the intersection of the average SE curves earlier.

3.4 Conclusion

In this chapter, we have proposed and studied a new discrete rate adaptive hierarchical modulation-assisted TBS scheme. The newly proposed scheme proposes to transmit information to single or two best users judiciously by the joint consideration of the channel qualities of the users. As such the frequency of information of the users increases without any degradation of overall SE compared with the classical single best user scheduling. For our newly proposed scheme we have derived expressions for the achievable overall SE and the information access probability. The presented numerical results, for an i.i.d. Rayleigh faded environment, have shown that our proposed scheme offers more frequent information access without any overall SE degradation compared with the conventional single best user scheduling.

Bibliography

- [36] Md. J. Hosssain, M.-S. Alouini, and V. K. Bhargava, "Multi-user opportunistic scheduling using power-controlled hierachical constellations," *IEEE Trans. Wireless Commun.*, to appear in May 2007.
- [37] P. K. Vitthaladevuni and M.-S. Alouini, "A recursive algorithm for the exact BER computation of generalized hierarchical QAM constellations," *IEEE Trans. Inform. Theory*, vol. 49, pp. 297-307, Jan. 2003.
- [38] A. J. Goldsmith and S.-G. Chua, "Variable-rate variable-power MQAM for fading channels," *IEEE Trans. Communs.*, vol. 45, pp. 1218-1230, Oct. 1997.
- [39] T. Eng, N. Kong, and L. B. Milstein, "Selection combining schemes for RAKE receivers," in *Proc. 4th Int. Conf. Univ. Personal Commun. (ICUPC'95)*, Tokyo, Japan, Nov. 1995, pp. 426-430.

Chapter 4

Two-User Opportunistic Scheduling using Hierarchical Modulations in Wireless Networks with Heterogenous Average Link Gains

⁷In [43], authors have analyzed various performances of *single user* selection diversity schemes in a multiuser environment. There are two important metrics in order to compare performance of opportunistic scheduling schemes. One is the degrees of fairness (DOF) which gives the measure of how the total resource, e.g., system's SE, is shared among the users. For a system with identical average fading gain, in the long run, all the users in the system have the fair sharing of the system's resource. However, DOF metric is very important for a multiuser system where different users have different average channel qualities. Another important metric is the frequency of information access which corresponds to the average wait time in order to get information access. This is important for the time-out consideration in the upper-layer protocols [43]. Even if the users have identical average channel fading gain, lower frequency of information access may result in higher queuing delay and packet loss rate for memory limited system [42].

In Chapter 3 we have proposed a rate adaptive hierarchical modulation-assisted TBS. The performance of the proposed TBS is also analyzed in a fading environment where all the users in the network experience *identical* average channel fading gain. In an wireless system, where cell size is not smaller and slow power control is used, identical average channel quality for different users can be achieved. However, for cellular system with

⁷The research work presented in this chapter is submitted as Md. J. Hossain, M.-S. Alouini, and V. K. Bhargava, "Two-User Opportunistic Scheduling using Hierarchical Modulations in Wireless Networks with Heterogenous Average Link Gains", submitted *IEEE Trans. Commun.*, May 2007.

larger cell size, different users may have different average channel quality due to the pathloss and shadowing differences. In this chapter, first, we analyze the performance of the rate adaptive hierarchical modulation-assisted TBS in a heterogenous fading environment where different users have non identical average link fading gains. Specifically, we present expression for the SE of the users and using this expression, we compare DOF of the TBS with that of classical single user opportunistic scheduling schemes, namely, absolute CNR-based SBS and normalized CNR based PFS. Second contribution is that we propose a new HTS scheme based on our earlier proposed TBS in Chapter 3 (also refer to [40]). This HTS selects the first user based on the largest absolute CNR value among all the users while the second user is selected based on the ratios of the absolute CNRs to the corresponding average CNRs of the remaining users. The total transmission rate i.e., the constellation size is selected according to the absolute CNR of the first best user. The total transmission rate is then allocated among these selected users by joint consideration of their absolute CNRs and allocated number of information bit(s) are transmitted to them using hierarchical modulations. Numerical results are presented for a fading environment where different users experience independent but non-identical (i.n.d.) channel fading.

4.1 System and Channel Models

We consider a single cell system where K users are communicating with a single BS. The BS transmits a fixed amount of power in each channel access unless a transmission outage is declared. It is assumed that the channel qualities of the users are perfectly estimated and reported to the BS without negligible delay in the form of absolute CNR. We consider that the fading amplitude of the users is characterized by the Rayleigh PDF which leads to the following PDF of kth user CNR g_k , $p_{g_k}(g)$

$$p_{g_k}(g) = \frac{1}{\bar{g}_k} \exp\left(-\frac{g}{\bar{g}_k}\right), \quad g_k \ge 0$$
(4.1)

where \bar{g}_k is the average channel gain of user k. Cumulative distribution function (CDF) of kth user's CNR, $P_{q_k}(x)$ is expressed as

$$P_{g_k}(x) = \int_0^x p_{g_k}(g) dg = \left(1 - \exp\left(-\frac{x}{\bar{g}_k}\right)\right). \tag{4.2}$$

4.2 Classical Single User Scheduling: Overview

4.2.1 Absolute CNR-based single best user scheduling

According to the classical opportunistic scheduling with uncoded adaptive $M_r = 2^{2r}$ -QAM constellations $(r \in \{1, 2, \dots, Q\})$, the scheduler at the BS selects the user (denoted by $U_{(1)}$) who has the largest CNR among all the users. If the CNR of the user $U_{(1)}$, $g_{(1)}$ falls in the *r*th fading region, the constellation size M_r is transmitted. The region boundaries $\{\gamma_r\}_{r=1}^{Q+1}$ are set to the CNR such that the information bits are sent below the target BER, BER₀ using the uniform M_r -QAM constellation. A transmission outage occurs if the $g_{(1)}$ falls below the threshold γ_1 . Since this scheme selects the user in a given time slot according to the absolute value of the instantaneous CNR, it is referred to as absolute CNR-based single user scheduling scheme (ASS) [43]

4.2.2 Normalized CNR-based single user scheduling

The PFS scheme selects a single user to be transmitted based on the relative channel strength [41] which is defined as the ratio of the data rate each individual's channel can afford to its average throughput. An alternate view point of the original PFS is presented in [44]. Basically, according to this view point, the BS transmits to user k with the largest ratio of the instantaneous CNR to its own short-term average CNR. This alternate view point is used in order to select the second best user in our newly proposed HTS scheme. Without loss of generality, let us consider mth user is the first best user at time slot n and its absolute CNR is $g_m(n)$. Then basically, the selected second user is the user who has the highest

$$\frac{g_k(t)}{\bar{g}_k}, \quad k \neq m, \tag{4.3}$$

where $g_k(n)$ and \bar{g}_k are kth user instantaneous CNR at time slot n. Once a user is selected to be transmitted, the constellation size M_r is transmitted if the instantaneous CNR of the selected user falls in the rth fading region. Since in each time slot, the BS selects the user with the largest CNR to its own average CNR, it has been referred to as the normalized CNR based scheduling scheme (NSS) in the literature. It offers the highest fairness among the users at the expense of certain amount of multiuser diversity gain.
4.3 Proposed Two-User Opportunistic Scheduling

Since we analyze the performance of the rate adaptive TBS in an i.n.d. fading environment and a HTS is proposed based on the TBS, in what follows, we briefly describe the hierarchical modulation-assisted TBS (refer to [40] for the detailed description of the TBS and its performance in an i.i.d fading environment). We then present our newly proposed HTS in Section 4.3.2.

4.3.1 Two-best user scheduling

The details of rate adaptive TBS scheme has been described in Chapter 3 (also refer to [40]). However, since in this chapter we propose a HTS capitalized on the concept of the rate adaptive hierarchical modulation-assisted TBS, we present a brief over view on this scheme. Basically, according to the available instantaneous CNR reporting from all the users, the BS scheduler selects two users, namely the *first best user* (denoted as $U_{(1)}$) and the *second best user* (denoted as $U_{(2)}$), whose channel gains are the first and second highest, respectively. Without loss of generality let us denote, the channel gains of user $U_{(1)}$ and $U_{(2)}$ by $g_{(1)}$ and $g_{(2)}$, respectively with $g_{(1)} \ge g_{(2)}$. The working principle of the rate adaptive hierarchical modulation-assisted TBS can be divided into two phases as described below.

Constellation size, M_r selection

The constellation size is selected according to the first best user CNR, $g_{(1)}$. Basically, the CNR range of the user $U_{(1)}$, $g_{(1)}$ is divided into Q + 1 fading regions; namely $r = 0, 1, 2, \dots, Q$. When the received CNR, $g_{(1)}$ is estimated to be in the *r*th fading region, the constellation size $M_r = 2^{2r}$ (where 2r is the total number of bits per M_r -QAM symbol) is transmitted. The region boundaries (or constellation size switching thresholds) $\{\gamma_r\}_{r=1}^{Q+1}$ are set to the CNRs such that the information bits are sent below the target BER, BER₀ using the uniform M_r -QAM constellation. Once the constellation size is selected, the total available transmission rate 2r is divided among the selected best users, $U_{(1)}$ and $U_{(2)}$ jointly considering their channel CNRs as described below.

Transmission rate allocation to the selected users $U_{(1)}$ and $U_{(2)}$

To transmit information to the selected second best user opportunistically, the scheme switches between the uniform and hierarchical constellations jointly considering the CNR estimate fluctuation of the two-best users as follows. As described in [40] that a QAM constellation of order $M_r = 2^{2r}$, $(r = 2, 3, \dots, Q)$ can form a set of hierarchical constellations $C_r = \{2^2/2^{2r}, 2^4/2^{2r} \dots, 2^{2(r-1)}/2^{2r}\}$. Each hierarchical constellation in the set, C_r has two different hierarchies, namely first and second hierarchies and it specifies a unique combinations of bits in two different hierarchies. Furthermore, these two different hierarchies have two different CNR requirements in order to decode the transmitted bits with a given reliability. Therefore, in a given fading region, the number of bits within the constellation that can be assigned to the users $U_{(1)}$ and $U_{(2)}$ jointly depends on the received CNR estimate of both the users. To assign the available transmission rate (i.e., the number of the bits) to the selected users, $U_{(1)}$ and $U_{(2)}$, the *r*th fading region $(r = 2, 3, \dots, Q)$ is divided into $s = 0, 1, \dots, r - 1$ sub-regions. The sub-region boundaries $\{\gamma_{r,s}^2\}_{r=2,s=1}^{Q,r-1}$ are obtained by numerically solving the BER equations of the hierarchical $2^{2s}/2^{2r}$ -QAM such that the bits assigned in the second hierarchy are sent below the target BER, BER₀.

When the CNR of the first best user is estimated in the *s*th sub-region in the *r*th fading region, 2(r - s) bits in the second hierarchy can be decoded by the first best users with the required reliability⁸. However, in order to decide whether to transmit information to the user $U_{(2)}$ or not, the channel CNR of the second best user, $g_{(2)}$ is checked. If $g_{(2)}$ is good enough to decode the 2s information bits transmitted in the first hierarchy with the required reliability, these bits are transmitted to the user $U_{(2)}$. Consequently, hierarchical $2^{2s}/2^{2r}$ -QAM is transmitted from the demodulator. Otherwise, the available total transmission rate r is assigned to the user $U_{(1)}$ using the uniform 2^{2r} -QAM. Therefore, if the channel CNR, $g_{(1)}$ falls in a given sub-region, the demodulator can transmit either the uniform constellation or the hierarchical constellation depending on the CNR estimate $g_{(2)}$. The switching between the uniform and the hierarchical constellations in a sub-region s ($s = 1, 2, \dots, r - 1$) of the fading region r ($r = 2, 3 \dots, Q$) depends on the CNR, $g_{(2)}$ and is determined by the thresholds, $\{\gamma_{r,s}^1\}_{r=2,s=1}^{Q,r-1}$. These thresholds $\gamma_{r,s}^1$ are selected such that the

⁸In sub-region 0 of a given region r corresponds to transmission to only the first best user using uniform 2^{2r} -QAM.

s bits assigned in the first hierarchy of $2^{2s}/2^{2r}$ -QAM are transmitted below the target BER, BER₀.

It is important to mention that in the first fading region (r = 1), the scheme proposes to transmit 4-QAM which has only one hierarchy. However, it has two different phases; inphase and quadrature phase. If the second best user's CNR can support 4-QAM as well, an information bit is transmitted in one of the phases for the first best user and an information bit is transmitted in the other phase for the second best user.

4.3.2 Newly proposed hybrid two-user scheduling

We will see in the numerical results section, in this chapter, that the TBS scheme can increase the system's fairness without any degradation of link SE compared with the classical SBS scheme. However, the system's fairness can further be improved if the second user is selected according to a selection criterion that is equivalent to the selection criterion of the PFS scheme. For the newly proposed HTS, the first user selection, the constellation size selection and the rate allocation procedures remain same as described in Section 4.3.1. The only difference is in the selection procedure of the second user. The proposed HTS scheme is described below in details.

- Step 1: Select the first best user, $U_{(1)}$ and its instantaneous CNR $g_{(1)}$, based on the absolute values of instantaneous CNRs of the users.
- Step 2: Remove user $U_{(1)}$ from the users' set. Then calculate \hat{g}_k using Eq. (4.3) for all k except user $U_{(1)}$ who has selected as the best first user in Step 1.
- Step 3: Find \hat{m} such that

$$\hat{m} = \arg \max \left(\hat{g}_k \right) \forall \quad k \text{ except user } \mathbf{U}_{(1)}$$

$$(4.4)$$

and set

$$g_{(2)} = g_{\hat{m}}.$$
 (4.5)

- Step 3: Select the constellation size, $M_r = 2^{2r}$ if $g_{(1)}$ falls in $[\gamma_r, \gamma_{r+1}]$.
- Step 4: Check whether $g_{(1)}$ falls in $[\gamma_{r,s}^2, \gamma_{r,s+1}^2]$ and $g_{(2)}$ falls in $[\gamma_{r,s}^1, \gamma_{r,s+1}^2]$. If it is true, transmit 2s bits of information for the user $\hat{U}_{(2)}$ and 2(r-s) bits of informa-

tion for the user $U_{(1)}$ using hierarchical $2^{2s}/2^{2r}$ -QAM constellation. Otherwise, 2r information bits are transmitted for user $U_{(1)}$ using M_r -QAM constellation.

It is important to mention that both users, in rate adaptive TBS scheme, are selected according to the *absolute* CNRs of the users. In contrast, the newly proposed HTS scheme selects the first user based on the largest value of absolute CNR of all the users whereas the second user is selected based on the ratio of absolute CNRs to the corresponding average CNRs of the users except the user who has been selected as the first user. The newly proposed two-user opportunistic scheduling is, therefore, referred to as the hybrid two-user scheduling.

4.4 **Performance Analysis**

4.4.1 Two-best user scheduling

In our following analysis, we consider *j*th user as the tagged user. Without loss of generality, we denote the first and the second largest CNR of all the users' CNRs except user *j*, by $g_{(1)}^*$ and $g_{(2)}^*$, respectively. It is important to mention that basically there are three possibilities, namely, Scenario 1, Scenario 2 and Scenario 3 of receiving information by user *j* with the rate adaptive TBS scheme. These scenarios and the associated probabilities are described below.

Scenario 1-Information transmission only to user j as the first best user: In such scenario, 2r information bits are transmitted only to jth user using uniform 2^{2r} -QAM constellation. This happens again in two cases. One is that if the CNR of jth user, g_j is greater or equal to the largest CNR of all other users, $g_{(1)}^*$ and it falls in the 0th sub-region of fading region r ($r = 2, 3 \cdots, Q$). The other is that the first largest CNR of all other users, $g_{(1)}^*$ is less than the threshold $\gamma_{r,s}^1$ although the CNR g_j is greater than or equal to $g_{(1)}^*$ and falls in sth sub-region ($s = 1, 2, \cdots, r - 1$) in the rth fading region (see Fig. 3.4 for details). Therefore, the total probability, $a_{r,0}^{j,1}$ that the uniform 2^{2r} -QAM constellation is transmitted to jth user in the rth fading region is the summation of all the individual probabilities and can be written as

$$a_{r,0}^{j,1} = \int_{\gamma_r}^{\gamma_{r,1}^2} p_{g_j}(x) \int_0^x p_{g_{(1)}^*}(z) dz dx + \sum_{s=1}^{r-1} \int_{\gamma_{r,s}^2}^{\gamma_{r,s+1}^2} p_{g_j}(x) \int_0^{\gamma_{r,s}^1} p_{g_{(1)}^*}(z) dz dx \qquad (4.6)$$

where $p_{g_{(1)}^{\star}}(x)$ is the PDF of the largest CNR of all the users' CNR except user j and which is expressed as [45]

$$p_{g_{(1)}^{\star}}(x) = \sum_{k=1, k \neq j}^{K} p_{g_k}(x) \prod_{m=1, m \neq k, j}^{K} P_{g_m}(x).$$
(4.7)

The first integral in Eq. (4.6) has the form of integral $I_1^j(t_1, t_2)$ presented in Eq. (9.4) of Appendix B whereas the second integral has the form of integral $I_2^j(t_1, t_2, t_3)$ presented in (9.6) of Appendix B. Therefore, $a_{r,0}^{j,1}$ in Eq. (4.6) can expressed as

$$a_{r,0}^{j,1} = I_1^j(\gamma_r, \gamma_{r,1}^2) + \sum_{s=1}^{r-1} I_2^j(\gamma_{r,s}^2, \gamma_{r,s+1}^2, \gamma_{r,s}^1)$$
(4.8)

where the integrals $I_1^j(\cdot, \cdot)$ and $I_1^j(\cdot, \cdot, \cdot)$ are evaluated in closed-form in Eqs. (9.5) and (9.7) of Appendix B, respectively.

Scenario 2-Information transmission to user j as the first best user along with the second best user: In such scenario, 2(r - s) information bits are transmitted to jth user (as the first best user) and 2s information bits are transmitted to another user (as the second best user) using hierarchical $2^{2s}/2^{2r}$ -QAM constellation. The probability, $a_{r,s}^{j,1}$ that the 2(r - s) bits of information are transmitted to jth user using $2^{2s}/2^{2r}$ -QAM constellation corresponds to the probability that the CNR, g_j is greater than $g_{(1)}^*$ and falls in sth subregion in the rth (r > 1) fading region while $g_{(1)}^*$ is above the threshold $\gamma_{r,s}^1$. Therefore, the probabilities $a_{r,s}^{j,1}$ can be written as

$$a_{r,s}^{j,1} = \begin{cases} \int_{\gamma_{r,s}^{2}}^{\gamma_{r,s+1}^{2}} p_{g_{j}}(x) \int_{\gamma_{r,s}^{1}}^{x} p_{g_{(1)}^{*}}(y) dy dx, & s = 1, 2, \cdots, r-2, \\ \\ \int_{\gamma_{r,s}^{2}}^{\gamma_{r+1}^{2}} p_{g_{j}}(x) \int_{\gamma_{r,s}^{1}}^{x} p_{g_{(1)}}(y) dy dx, & s = r-1. \end{cases}$$

$$(4.9)$$

Both integrals in Eq. (4.9) have the form of integral $I_3^j(t_1, t_2, t_3)$ which is evaluated in closed-form in Eq. (9.9) of Appendix B. Therefore, the probabilities $a_{r,s}^{j,1}$ can be expressed as

$$a_{r,s}^{j,1} = \begin{cases} I_3^j(\gamma_{r,s}^2, \gamma_{r,s+1}^2, \gamma_{r,s}^1) & s = 1, \cdots, r-2, \\ I_3^j(\gamma_{r,s}^2, \gamma_{r+1}^2, \gamma_{r,s}^1) & s = r-1. \end{cases}$$
(4.10)

Recall from Section 4.3.1 that in fading region r = 1, 4-QAM constellation is transmitted and the first best user receives information along with the second best user if the second best user's channel quality is above the threshold γ_1 . In such scenario, 1 bit of information is transmitted in the inphase channel for the first best user and 1 bit of information is transmitted for the second best user in other phase. The probability, $a_{1,1}^{j,1}$ that 1 bit of information is transmitted to *j*th user using uniform 4-QAM constellation while it is the first best user, can be written as

$$a_{1,1}^{j,1} = \int_{\gamma_1}^{\gamma_2} p_{g_j}(x) \int_{\gamma_1}^x p_{g_{(1)}^*}(y) dy dx.$$
(4.11)

The integral in Eq. (4.11) has the form of integral $I_3^j(t_1, t_2, t_3)$ in Eq. (9.8) which is given in closed-form in Eq. (9.9) of Appendix B. Therefore, $a_{11}^{j,1}$ can be expressed as

$$a_{1,1}^{j,1} = I_3^j(\gamma_1, \gamma_2, \gamma_1). \tag{4.12}$$

Scenario 3-Information transmission to user j as the second best user along with the first best user: In this scenario, jth user receives 2s bits of information as the second best user along with the another user (as the first best user) using hierarchical $2^{2s}/2^{2r}$ -QAM constellation. This happens if g_j is equal to or greater than $g_{(2)}^*$ as well as it falls above $\gamma_{r,s}^1$ while the CNR $g_{(1)}^*$ falls in the sth subregion ($s = 1, 2, \dots, r-1$) of the rth fading region. This probability, $a_{r,s}^{j,2}$ can be written as

$$a_{r,s}^{j,2} = \begin{cases} \int_{\gamma_{r,s}^{2}}^{\gamma_{r,s+1}^{2}} \int_{\gamma_{r,s}^{1}}^{x} p_{g_{j}}(y) \int_{0}^{y} p_{g_{(1)}^{*}g_{(2)}^{*}}(x,z) dz dy dx, & s = 1, 2, \cdots, r-2, \\ \\ \int_{\gamma_{r,s}^{2}}^{\gamma_{r+1}^{2}} \int_{\gamma_{r,s}^{1}}^{x} p_{g_{j}}(y) \int_{0}^{y} p_{g_{(1)}^{*}g_{(2)}^{*}}(x,z) dz dy dx. & s = r-1 \end{cases}$$

$$(4.13)$$

where $p_{g_{(1)}^*g_{(2)}^*}(x, z)$ is the joint PDF of the largest two CNRs of all the users' CNRs except user j and which is given by [45]

$$p_{g_{(1)}^{*}g_{(2)}^{*}}(x,z) = \sum_{k,k\neq j}^{K} p_{g_{k}}(x) \sum_{l=1,l\neq j,k}^{K} p_{g_{l}}(z) \prod_{m=1,m\neq j,k,l}^{K} P_{g_{m}}(z).$$
(4.14)

The integrals in Eq. (4.13) are of the form of integral $I_4^j(t_1, t_2, t_3)$ in Eq. (9.10) of Appendix B. Therefore, $a_{r,s}^{j,2}$ can be expressed as

$$a_{r,s}^{j,2} = \begin{cases} I_4^j(\gamma_{r,s}^2, \gamma_{r,s+1}^2, \gamma_{r,s}^1) & s = 1, \cdots, r-2, \\ I_4^j(\gamma_{r,s}^2, \gamma_{r+1}^2, \gamma_{r,s}^1) & s = r-1 \end{cases}$$
(4.15)

where $I_4^j(\cdot, \cdot, \cdot)$ is evaluated in closed-form in Eq. (9.11) of Appendix B.

The probability, $a_{1,1}^{j,2}$ that 1 bit of information is transmitted to *j*th user using the uniform 4-QAM constellation while it is the second best user can be expressed as

$$a_{1,1}^{j,2} = \int_{\gamma_1}^{\gamma_2} \int_{\gamma_1}^x p_{g_j}(y) \int_0^y p_{g_{(1)}^*g_{(2)}^*}(x,z) dz dy dx.$$
(4.16)

This integral has the form of integral $I_4^j(t_1, t_2, t_3)$ in Eq. (9.10) of Appendix B. Therefore, $a_{11}^{j,2}$ can be expressed as

$$a_{1,1}^{j,2} = I_4^j(\gamma_1, \gamma_2, \gamma_1). \tag{4.17}$$

Spectral efficiency

The average link spectral efficiencies of *j*th user as the first and the second best user, respectively, $\langle R_{(i)}^j \rangle / W$ (i = 1, 2) are just the sum of data rates associated with the individual sub-regions of each region, weighted by the probabilities $a_{r,s}^{j,1}$ and $a_{r,s}^{j,2}$, respectively and can be expressed as

$$\frac{\langle R_{(1)}^{j} \rangle}{W} = \sum_{r=1}^{Q} \sum_{s=0}^{r-1} 2(r-s) a_{r,s}^{j,1} + a_{1,1}^{j,1}, \qquad (4.18)$$

$$\frac{\langle R_{(2)}^j \rangle}{W} = \sum_{r=2}^Q \sum_{s=1}^{r-1} 2s \; a_{r,s}^{j,2} + a_{1,1}^{j,2}. \tag{4.19}$$

Since a given user can get information access in a given time slot either being as the first best or as the second best user, the overall SE of *j*th user with TBS, $\frac{\langle R_{\text{TBS}}^j \rangle}{W}$ is the summation of individual SE as the first and second best users i.e.,

$$\frac{\langle R_{\text{TBS}}^j \rangle}{W} = \frac{\langle R_{(1)}^j \rangle}{W} + \frac{\langle R_{(2)}^j \rangle}{W}.$$
(4.20)

Using Eqs. (4.18) and (4.19), $\frac{\langle R_{\text{TBS}}^j \rangle}{W}$ in Eq. (4.20) can be written as

$$\frac{\langle R_{\text{TBS}}^j \rangle}{W} = \sum_{r=1}^Q \sum_{s=0}^{r-1} 2(r-s) a_{r,s}^{j,1} + a_{1,1}^{j,1} + \sum_{r=2}^Q \sum_{s=1}^{r-1} 2s a_{r,s}^{j,2} + a_{1,1}^{j,2}.$$
(4.21)

The system's link SE with TBS, $\frac{\langle R_{\text{TBS}} \rangle}{W}$ can be obtained by adding users' individual spectral efficiencies as follows:

$$\frac{\langle R_{\rm TBS} \rangle}{W} = \sum_{j=1}^{K} \frac{\langle R_{\rm TBS}^j \rangle}{W}.$$
(4.22)

Using Eq. (4.21) in Eq. (4.20), $\frac{\langle R_{\text{TBS}} \rangle}{W}$ can expressed as

$$\frac{\langle R_{\text{TBS}} \rangle}{W} = \sum_{j=1}^{K} \left(\sum_{r=1}^{Q} \sum_{s=0}^{r-1} 2(r-s) a_{r,s}^{j,1} + a_{1,1}^{j,1} + \sum_{r=2}^{Q} \sum_{s=1}^{r-1} 2s a_{r,s}^{j,2} + a_{1,1}^{j,2} \right).$$
(4.23)

Degree of fairness

There are different ways to quantify the DOF (see [43] and references there in). In this dissertation, the DOF for *j*th user is measured as the portion of the system's link SE is attributed by *j*th user. Therefore, the DOF of user j, F^{j} , in general, is defined as

$$F^{j} = \frac{-\log\left(\frac{\langle R^{j} \rangle / W}{\langle R \rangle / W}\right)}{-\log(\frac{1}{K})}$$
(4.24)

where $\langle R^j \rangle / W$ is the *j*th user's SE, $\langle R \rangle / W$ is the system's link SE. Therefore, F^j is the proportion of system's resource (e.g., the link SE) attributed to user *j*. In general, the average fairness of a multiuser system is then defined as

$$\bar{F} = -\sum_{k=1}^{K} \frac{\langle R^k \rangle / W}{\langle R \rangle / W} \times \frac{\log\left(\frac{\langle R^k \rangle / W}{\langle R \rangle / W}\right)}{\log(K)}.$$
(4.25)

Therefore, the average fairness of a multiuser system with TBS scheme, \bar{F}_{TBS} can be written as

$$\bar{F}_{\text{TBS}} = -\sum_{k=1}^{K} \frac{\langle R_{\text{TBS}}^k \rangle / W}{\langle R_{\text{TBS}} \rangle / W} \times \frac{\log\left(\frac{\langle R_{\text{TBS}}^k \rangle / W}{\langle R_{\text{TBS}} \rangle / W}\right)}{\log(K)}.$$
(4.26)

 $\overline{F}_{\text{TBS}}$ in Eq. (4.26) can be evaluated using Eqs. (4.21) and (4.23).

4.4.2 Hybrid two-user scheduling

In order to evaluate the SE performance for our proposed rate adaptive HTS, we need the expression for the joint PDF of the channel fading gains of the first best user and second user who is selected according to the normalized CNR. Once the expression of the joint PDF is available, the SE well as the DOF performances of the HTS can be found analytically using similar methodology as described above. However, to the best of our knowledge, the expression for the joint PDF is not known. Finding this expression remians as open research problem. We evaluate the performance of the proposed HTS via computer simulation as reported in Section 4.5.

4.4.3 Single user scheduling

Spectral efficiency

In absolute CNR-based SBS, the SE of *j*th user and the system's link SE with adaptive M-QAM transmission, $\frac{\langle R_{ASS}^j \rangle}{W}$ can be written as

$$\frac{\langle R_{\text{ASS}}^j \rangle}{W} = \sum_{r=0}^Q 2r \ a_r^j \tag{4.27}$$

where a_r^j are the probabilities that *j*th user's CNR is larger than the largest CNR of all other users' CNR and falls in the *r*th fading region. Therefore, a_r^j can be expressed as

$$a_r^j = \int_{\gamma_r}^{\gamma_{r+1}} p_{g_j}(x) \int_0^x p_{g_1^*}(y) dy dx.$$
(4.28)

Using Eqs. (4.1) and (4.7), a_r^j can be evaluated as

$$a_{r}^{j} = \int_{\gamma_{r}}^{\gamma_{r+1}} p_{g_{j}}(x) \prod_{k=1, k \neq j}^{K} \left(1 - \exp\left(-\frac{x}{g_{k}}\right)\right) dx$$

$$= \int_{\gamma_{r}}^{\gamma_{r+1}} p_{\gamma_{j}}(x) \sum_{\tau' \in T_{j}^{K}} \operatorname{sign}(\tau') \exp\left(-\tau'x\right) dx$$

$$= \int_{\gamma_{r}}^{\gamma_{r+1}} \sum_{\tau' \in T_{j}^{K}} \frac{\operatorname{sign}(\tau')}{\bar{g}_{j}} \exp\left(-x(\tau' + \frac{1}{\bar{g}_{j}})\right) dx$$

$$= \sum_{\tau' \in T_{j}^{K}} \frac{\operatorname{sign}(\tau')}{\tau' \bar{g}_{j} + 1} \left[\exp\left(-\frac{\gamma_{r}(\bar{g}_{j}\tau' + 1)}{\bar{g}_{j}}\right) - \exp\left(-\frac{\gamma_{r+1}(\bar{g}_{j}\tau' + 1)}{\bar{g}_{j}}\right)\right].(4.29)$$

Using Eq. (4.29) in Eq. (4.27), average SE of *j*th user, $\frac{\langle R_{ASS}^j \rangle}{W}$ can finally be expressed as

$$\frac{\langle R_{\text{ASS}}^j \rangle}{W} = \sum_{r=0}^Q \sum_{\tau' \in T_j^K} \frac{2r \text{sign}(\tau')}{(\tau' \bar{g}_j + 1)} \left[\exp\left(-\frac{\gamma_r(\bar{g}_j \tau' + 1)}{\bar{g}_j}\right) - \exp\left(-\frac{\gamma_{r+1}(\bar{g}_j \tau' + 1)}{\bar{g}_j}\right) \right].$$
(4.30)

The system's link SE with AAS scheme, $\frac{\langle R_{ASS} \rangle}{W}$ is the summation of individual user's SE and can be expressed in desired closed-form as follows:

$$\frac{\langle R_{\text{ASS}} \rangle}{W} = \sum_{j=1}^{K} \sum_{r=0}^{Q} \sum_{\tau' \in T_j^K} \frac{2r \text{sign}(\tau')}{(\tau' \bar{g}_j + 1)} \left[\exp\left(-\frac{\gamma_r(\bar{g}_j \tau' + 1)}{\bar{g}_j}\right) - \exp\left(-\frac{\gamma_{r+1}(\bar{g}_j \tau' + 1)}{\bar{g}_j}\right) \right].$$

$$(4.31)$$

In NSS scheme, BS selects user j if it has the largest normalized CNR, i.e., [43]

$$\eta_j = \frac{g_j}{\bar{g}_j}.\tag{4.32}$$

and the PDF of η_j is given by

$$p_{\eta_j}(x) = \exp(-x).$$
 (4.33)

Therefore, in NSS scheme, the SE of *j*th user, $\frac{\langle R_{NSS}^j \rangle}{W}$ can be obtained as

$$\frac{\langle R_{\text{NSS}}^j \rangle}{W} = \sum_{r=0}^Q 2r \ b_r^j \tag{4.34}$$

where b_r^j are the probabilities that the *j*th user normalized CNR is larger than the maximum normalized CNR of all other users' normalized CNR and its absolute CNR falls in the *r*th fading region. Equivalently, b_r^j can be expressed as

$$b_r^j = \int_{\hat{\gamma}_r}^{\hat{\gamma}_{r+1}} p_{\eta_j}(x) \int_0^x p_{\eta_1^*}(y) dy dx$$
(4.35)

where $\{\hat{\gamma}_r\}_{r=0}^{r=Q+1}$ are the values of constellation size switching thresholds normalized by the average CNR of *j*th user i.e., $\hat{\gamma}_r = \gamma_r/\bar{g}_j$. Using Eqs. (4.33) and (4.7), b_r^j can be evaluated as

$$b_{r}^{j} = \int_{\hat{\gamma}_{r}}^{\hat{\gamma}_{r+1}} p_{\eta_{j}}(x) \prod_{k=1, k \neq j}^{K} (1 - \exp(-x)) dx$$

= $\int_{\hat{\gamma}_{r}}^{\hat{\gamma}_{r+1}} \exp(-x) (1 - \exp(-x))^{K-1} dx$
= $\frac{1}{K} \left[1 - \left(\exp\left(-\frac{\gamma_{r}}{\bar{g}_{j}}\right) - \exp\left(-\frac{\gamma_{r+1}}{\bar{g}_{j}}\right) \right)^{K} \right].$ (4.36)

Using the Eq. (4.36) in Eq. (4.34), average SE of *j*th user, $\frac{\langle R_{NSS}^j \rangle}{W}$ can be written in simplified form as

$$\frac{\langle R_{\text{NSS}}^j \rangle}{W} = \sum_{r=0}^{Q} \frac{2r}{K} \left[1 - \left(\exp\left(-\frac{\gamma_r}{\bar{g}_j}\right) - \exp\left(-\frac{\gamma_{r+1}}{\bar{g}_j}\right) \right)^K \right].$$
(4.37)

The system's link SE with NSS scheme, $\frac{\langle R_{NSS}^j \rangle}{W}$ can finally be written as

$$\frac{\langle R_{\text{NSS}} \rangle}{W} = \sum_{j=1}^{K} \sum_{r=0}^{Q} \frac{2r}{K} \left[1 - \left(\exp\left(-\frac{\gamma_r}{\bar{g}_j}\right) - \exp\left(-\frac{\gamma_{r+1}}{\bar{g}_j}\right) \right)^K \right].$$
(4.38)

Degree of fairness

The DOF of user j with ASS scheme, F_{ASS}^{j} can be written as

$$\bar{F}_{ASS} = -\sum_{k=1}^{K} \frac{\langle R_{ASS}^{k} \rangle / W}{\langle R_{ASS} \rangle / W} \times \frac{\log\left(\frac{\langle R_{ASS}^{k} \rangle / W}{\langle R_{ASS} \rangle / W}\right)}{\log(K)}.$$
(4.39)

 \overline{F}_{ASS} in Eq. (4.39) can be evaluated using Eqs. (4.30) and (4.31).

Similarly, the DOF of a multiuser system with NSS, \mathcal{F}_{ASS}^{j} can be expressed as

$$\bar{F}_{\text{NSS}} = -\sum_{k=1}^{K} \frac{\langle R_{\text{NSS}}^k \rangle / W}{\langle R_{\text{NSS}} \rangle / W} \times \frac{\log\left(\frac{\langle R_{\text{NSS}}^k \rangle / W}{\langle R_{\text{NSS}} \rangle / W}\right)}{\log(K)}$$
(4.40)

which can be evaluated using Eqs. (4.37) and (4.38).

4.5 Numerical Results

We assume a target BER of 10^{-4} and a maximum M-QAM constellation size of 1024. Further, we assume a 10 users system where different users are located in 10 different tiers inside the cell. The tiers are uniformly distributed. The average link gain of the far-end user which is located in the cell boundary is assumed to be 10[dB]. We consider that kth user is located in tier k from the cell boundary and its average CNR, $\bar{g}_k = (10 + \Delta \bar{g} \times k) [dB]$ where $\Delta \bar{q}$ [dB] is the deviation of average link gains of two successive users due to the pathloss. The near-end user is the closest user to the BS and it has the highest average channel fading gain. For this set up, the system's DOF is plotted in Fig. 4.1 for various values of deviation of average link gains of near-end and far-end users by varying the value of $\Delta \bar{g}$. Four different scheduling schemes are considered in this figure. This figure shows that the newly proposed HTS scheme considerably increases the system's fairness in comparison to the classical absolute CNR-based SBS. Interestingly, there is no degradation of the system SE in comparison to the classical SBS as seen from Fig. 4.2. The reason is that HTS selects the constellation size according to the largest instantaneous CNR fluctuations and the second user is included in a transmission judiciously using the hierarchical modulations. It is also important to mention that since the newly proposed HTS selects the second user based on the normalized CNR criteria, it achieves better fairness than the TBS scheme. On the other hand, the NSS schme i.e., the equivalent PFS, has the highest system's fairness as



Figure 4.1: Average system's fairness versus number of users.



Figure 4.2: Link spectral efficiency versus number of users (link SE curves for TBS, SBS and HTS overlap).

observed in Fig. 4.1 but this fairness comes at the expense of a certain amount of link SE as observed in Fig. 4.2.

4.6 Conclusion

In this chapter, we have analyzed the performance of the hierarchical modulation-assisted TBS in a fading environment where different users have different average link qualities. Specifically, we have derived the expressions for the SE, and the DOF of the users. Using these expressions we have compared the DOF of our earlier proposed rate adaptive TBS scheme with those of classical single user opportunistic scheduling schemes. We have also proposed a new HTS scheme. Some selected numerical results showed that the TBS

offers higher DOF in a multiuser system with heterogenous link gains than the classical SBS scheme without any degradation of system SE. The HTS scheme offers higher DOF than both TBS and SBS without any degradation of system SE. Compared with the NSS scheme, both TBS and HTS has lower DOF but NSS suffers from certain degradation of system SE.

Bibliography

- [40] Md. J. Hosssain, M.-S. Alouini, and V. K. Bhargava, "Rate adaptive hierarchical modulation-assisted two-user opportunistic scheduling," *IEEE Trans. Wireless Commun.*, to appear in Jun. 2007.
- [41] D. N. Tse, P. Viswanath, and R. Laroia, "Opportunistic beamforming using dumb antennas," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1277-1294, Jun. 2002.
- [42] Md. J. Hosssain, M.-S. Alouini, and V. K. Bhargava, "Upper-layer performance analysis of multiple best user opportunistic scheduling scheme using power-controlled hierarchical constellations," in *Proc. of IEEE Global Telecommun. Conf. (Globecom'06)*, pp. 1267-1272, San Fransisco, CA, Nov. 2006, pp.
- [43] L. Yang, and M.-S. Alouini, "Performance analysis of Multiuser Selection Diversity," *IEEE Trans. Veh. Technol.*, vol. 55, pp. 1003-1018, May 2006.
- [44] F. Berggren and R. Jantti, "Asymptotically fair transmission schedulign over fading channels," *IEEE Trans. Wireless Commun.*, vol. 3, pp. 326-336, Jan. 2004.
- [45] T. Eng, N. Kong, and L. B. Milstein, "Selection combining schemes for RAKE receivers," in *Proc. 4th Int. Conf. Univ. Personal Commun. (ICUPC'95)*, Tokyo, Japan, Nov. 1995, pp. 426-430.

Chapter 5

Queuing Delay and Buffer Distribution of Two-User Opportunistic Scheduling Schemes in Wireless Networks

⁹As we mentioned in Chapter 1 that serviced users in a wireless network with classical SBS scheme suffers from low frequency of information access. This results in higher queuing delay of the packets at the transmitter buffer. This has motivated us to design two-user opportunistic scheduling schemes which are presented in Chapters 2, 3 and 4. In Chapter 2, we also analyzed some link layer performances of our proposed power-controlled hierarchical modulation-assisted multi-user opportunistic scheduling scheme. Specifically, we analyzed and compared buffer distribution, average buffer occupancy (or delay) and PLP of both SBS scheme and power-controlled hierarchical modulation-assisted TBS via simple queuing analysis. Recently authors in [47] have analyzed various delay statistics of the classical SBS scheme via queuing analysis. The importance of delay analysis is to facilitate an admission controller design in order to maintain a statistical guarantee of delay bound. On the other hand, buffer distribution provides a picture of the average buffer occupancy level and consequently, the average delay. Inspired by the work in [47], we analyze in this chapter, buffer and delay distributions of our proposed rate adaptive hierarchical modulation-assisted TBS and HTS schemes. Then we compare these distributions with those of the classical single user scheduling schemes namely, SBS and NSS.

⁹The research work presented in this chapter is submitted as Md. J. Hossain, M.-S. Alouini, and V. K. Bhargava, "Queuing Delay and Buffer Distribution of Two-User Opportunistic Scheduling Schemes in Wireless Networks", to be submitted in *IEEE Trans. Wireless Commun.*, June 2007.

5.1 Overall System Description and Two-User Scheduling Schemes

Let us consider a BS that is transmitting to K mobile users opportunistically. We assume that the channel fading gain among the users varies independently. The users have a target PER requirement and a corresponding BER. It is also assumed that the channel fading of a given user varies independently from one time slot to the next time slot. For these K users, there are K separate radio link level buffers that are located at the BS. Data packets are arriving from upper-layer to these buffers with a Bernouli arrival process¹⁰. A scheduler is deployed at the BS to schedule the transmissions corresponding to different users in a TDM fashion. Transmissions occur within the fixed-sized time slots. Once a user is selected to be transmitted with a granted transmission rate, the granted number of packets are taken out from selected user's BS buffer and transmitted over the downlink to the selected user. With classical SBS, only one user is granted to receive information in a given time slot. With rate adaptive modulation assisted two user scheduling schemes, one or two users are granted to receive information in a given time slot depending on the channel qualities of the selected users. The details of the rate adaptive modulation-assisted TBS and HTS schemes were described in Chapter 3 and 4, respectively. However, we provide a brief description on each scheme.

5.1.1 Rate adaptive hierarchical modulation-assisted TBS scheme

Basically, according to the available instantaneous CNR reporting from all the users, the BS scheduler selects two users, namely the *first best user* (denoted as $U_{(1)}$) and the *second best user* (denoted as $U_{(2)}$), whose channel gains are the first and second highest, respectively. Without loss of generality let us denote, the channel gains of user $U_{(1)}$ and $U_{(2)}$ by $g_{(1)}$ and $g_{(2)}$, respectively with $g_{(1)} \ge g_{(2)}$. The working principle of the rate adaptive hierarchical modulation-assisted TBS can be divided into two phases as described below.

¹⁰For the sake of simplicity, we use Bernouli arrival process. However, we also provide simulation results for Markov modulated Poisson (MMP) arrival process.

Constellation size, M_r selection

The constellation size is selected according to the first best user CNR, $g_{(1)}$. Basically, the CNR range of the user $U_{(1)}$, $g_{(1)}$ is divided into R + 1 fading regions; namely $r = 0, 1, 2, \dots, Q$. When the received CNR, $g_{(1)}$ is estimated to be in the *r*th fading region, the constellation size $M_r = 2^{2r}$ (where 2r is the total number of bits per M_r -QAM symbol) is transmitted. The region boundaries (or constellation size switching thresholds) $\{\gamma_r\}_{r=1}^{Q+1}$ are set to the CNRs such that the information bits are sent below the target BER, BER₀ using the uniform M_r -QAM constellation. Once the constellation size is selected, the total available transmission rate 2r is divided among the selected best users, $U_{(1)}$ and $U_{(2)}$ jointly considering their channel CNRs as described below.

Transmission rate allocation to the selected users $U_{(1)}$ and $U_{(2)}$

To transmit information to the selected second best user opportunistically, the scheme switches between the uniform and hierarchical constellations jointly considering the CNR estimate fluctuation of the two-best users as follows. As described in [46], a QAM constellation of order $M_r = 2^{2r}$, $(r = 2, 3, \dots, Q)$ can form a set of hierarchical constellations $C_r = \{2^2/2^{2r}, 2^4/2^{2r} \dots, 2^{2(r-1)}/2^{2r}\}$. Each hierarchical constellation in the set, C_r has two different hierarchies, namely first and second hierarchies and it specifies a unique combinations of bits in two different hierarchies. Furthermore, these two different hierarchies have two different CNR requirements in order to decode the transmitted bits with a given reliability. Therefore, in a given fading region, the number of bits within the constellation that can be assigned to the users $U_{(1)}$ and $U_{(2)}$ jointly depends on the received CNR estimate of both users. To assign the available transmission rate (i.e., the number of the bits) to the selected users, $U_{(1)}$ and $U_{(2)}$, the *r*th fading region $(r = 2, 3, \dots, Q)$ is divided into $s = 0, 1, \dots, r - 1$ sub-regions. The sub-region boundaries $\{\gamma_{r,s}^2\}_{r=2,s=1}^{Q,r-1}$ are obtained by numerically solving the BER equations of the hierarchical $2^{2s}/2^{2r}$ -QAM such that the bits assigned in the second hierarchy are sent below the target BER, BER_0.

When the CNR of the first best user is estimated in the *s*th sub-region in the *r*th fading region, 2(r - s) bits in the second hierarchy can be decoded by the first best users with the required reliability¹¹. However, in order to decide whether to transmit information to the

¹¹In sub-region 0 of a given region r corresponds to transmission to only the first best user using uniform

user $U_{(2)}$ or not, the channel CNR of the second best user, $g_{(2)}$ is checked. If $g_{(2)}$ is good enough to decode the 2s information bits transmitted in the first hierarchy with the required reliability, these bits are transmitted to the user $U_{(2)}$. Consequently, hierarchical $2^{2s}/2^{2r}$ -QAM is transmitted from the demodulator. Otherwise, the available total transmission rate r is assigned only to user $U_{(1)}$ using the uniform 2^{2r} -QAM. Therefore, if the channel CNR, $g_{(1)}$ falls in a given sub-region, the demodulator can transmit either the uniform constellation or the hierarchical constellation depending on the CNR estimate $g_{(2)}$. The switching between the uniform and the hierarchical constellations in a sub-region s (s = $1, 2, \dots, r - 1$) of the fading region r ($r = 2, 3 \dots, Q$) depends on the CNR, $g_{(2)}$ and is determined by the thresholds, $\{\gamma_{r,s}^1\}_{r=2,s=1}^{Q,r-1}$. These thresholds $\gamma_{r,s}^1$ are selected such that the s bits assigned in the first hierarchy of $2^{2s}/2^{2r}$ -QAM are transmitted below the target BER, BER₀.

It is important to mention that in the first fading region (r = 1), the scheme proposes to transmit 4-QAM which has only one hierarchy. However, it has two different phases; inphase and quadrature phase. If the second best user's CNR can support 4-QAM as well, an information bit is transmitted in one of the phases for the first best user and an information bit is transmitted in the other phase for the second best user.

5.1.2 Rate adaptive hierarchical modulation-assisted HTS

The details of the proposed HTS scheme is described in what follows:

- Step 1: Select the first best user, $U_{(1)}$ and its instantaneous CNR $g_{(1)}$, based on the absolute values of instantaneous CNRs of the users.
- Step 2: Remove user $U_{(1)}$ from the users' set. Then calculate \hat{g}_k using Eq. (4.3) for all k except user $U_{(1)}$ who has selected as the best first user in Step 1.
- Step 3: Find \hat{m} such that

$$\hat{m} = \arg \max \left(\hat{g}_k \right) \forall \quad k \quad \text{except user } \mathbf{U}_{(1)}$$
 (5.1)

and set

$$g_{(2)} = g_{\hat{m}}.$$
 (5.2)

 2^{2r} -QAM.

- Step 3: Select the constellation size, $M_r = 2^{2r}$ if $g_{(1)}$ falls in $[\gamma_r, \gamma_{r+1}]$.
- Step 4: Check whether $g_{(1)}$ falls in $[\gamma_{r,s}^2, \gamma_{r,s+1}^2]$ and $g_{(2)}$ falls in $[\gamma_{r,s}^1, \gamma_{r,s+1}^2]$. If it is true, transmit 2s bits of information for the user $\hat{U}_{(2)}$ and 2(r-s) bits of information for the user $U_{(1)}$ using hierarchical $2^{2s}/2^{2r}$ -QAM constellation. Otherwise, 2r information bits are transmitted for user $U_{(1)}$ using M_r -QAM constellation.

5.2 Formulation of the Queueing Model

Let random variable r_n^j represent the number of packets that are granted to transmit to user j at time slot n. We assume that random variable b_n^j represents the number packets at jth buffer at time slot n and a packet arriving during time interval n - 1 cannot be transmitted until time interval n. If, in a given time slot, n buffer of jth user has less number of packets than the granted transmission rate r_n^j , min (b_n^j, r_n^j) packets are transmitted from jth buffer during that transmission slot. Traffic arrival probability of Bernouli arrival process at the jth buffer process is assumed to be μ^j . Without loss of generality, we consider the jth user's buffer as the target queue for queuing analysis and drop the superscript for user number in our formulation presented in the next sections. Assuming an infinite buffer size, the state space of the jth user's buffer occupancy and its buffer serving process can be written as follows:

$$\Omega = (b_n, s_n); \ b_n \ge 0; \ s_n \in \{0, 1\}.$$

5.2.1 Markov chain analysis

We model the system as a *discrete time Markov chain* (DTMC) whose transition matrix can be derived in terms of the state space above by considering the system dynamics. In order to find the transition probability matrix for the joint state of *j*th buffer, we derive the state transition matrix for service process r_n , **M** as follows:

$$\mathbf{M} = \begin{bmatrix} \mathbf{r}_{0 \to 0} & \mathbf{r}_{0 \to 1} & \mathbf{r}_{0 \to 2} & \mathbf{r}_{0 \to 3} & \cdots & \mathbf{r}_{0 \to 2Q} \\ \mathbf{r}_{1 \to 0} & \mathbf{r}_{1 \to 1} & \mathbf{r}_{1 \to 2} & \mathbf{r}_{1 \to 3} & \cdots & \mathbf{r}_{1 \to 2Q} \\ \mathbf{r}_{2 \to 0} & \mathbf{r}_{2 \to 1} & \mathbf{r}_{2 \to 2} & \mathbf{r}_{2 \to 3} & \cdots & \mathbf{r}_{2 \to 2Q} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{r}_{2Q \to 0} & \mathbf{r}_{2Q \to 1} & \mathbf{r}_{2Q \to 2} & \mathbf{r}_{2Q \to 3} & \cdots & \mathbf{r}_{2Q \to 2Q} \end{bmatrix},$$
(5.3)

where $r_{l \rightarrow k}$ is the probability that the serving rate of target buffer transits from rate *l* to *k* and 2*Q* is the possible maximum transmission rate. These transition probabilities depend on the scheduling policy that are employed at the BS. In fact, the presented analysis can be used for any scheduling policy once the transition matrix in Eq. (5.3) is known for that policy. In this chapter our objective is to analyze the buffer distribution and CDF of queuing delay of our proposed rate adaptive hierarchical modulation-assisted TBS as well as the HTS schemes and compare their performances with those of the classical single user scheduling schemes. Therefore, in what follows, we describe the service rate process and its transition probabilities for our proposed rate adaptive TBS scheme.

As it was described in Section 5.1.1, we use square QAM constellation of sizes 2^{2r} $(r = 1, 2, \dots, Q)$ for our proposed rate adprive TBS scheme. The total transmission rate, therefore, in a given time slot belongs to the set $\{2, 4, 6, \dots, Q\}$. With the rate adaptive TBS the total transmission rate is allocated among the selected users according to their channel qualities. Recall from Section 5.1.1 that when both selected users' channel qualities can support 4-QAM constellation, one packet is transmitted for each of the selected users. Therefore, the transmission rate with our rate adaptive TBS scheme, r_n belongs to the set $\mathcal{R} = \{0, 1, 2, 4, 6, \dots, 2Q\}$. Assuming that users' channels vary independently from one time slot to the next time slot, the transition probability, $r_{l\to k}$ corresponds to the probability that l packets are transmitted from jth buffer in a given time slot. This probability, $\operatorname{Prob}(r_n = l)$ can be written as

$$\operatorname{Prob.}(r_n = l) = \begin{cases} a_{1,1}^{j,1} + a_{1,1}^{j,2}, & \text{if } l = 1, \\ \sum_{r=\frac{l}{2}}^{Q} \sum_{s=0}^{r-l/2} a_{r,s}^{j,1} + \sum_{r=l/2+1}^{Q} a_{r,1}^{j,2}, & \text{if } l \neq 0, 1 \end{cases}$$
(5.4)

where $a_{r,s}^{j,1}$ denote the probabilities that 2(r-s) information packets are transmitted to *j*th user that is selected as the first best user. Similarly $a_{r,s}^{j,2}$ denote the probabilities that 2s packets are transmitted to *j*th user that is selected as the second best user. These probabilities have been evaluated in Eqs. (4.10), (4.12), (4.14) and (4.16) of Chapter 4. Please refer to these equations and there after for further details. The probability that no packet is

served from *j*th buffer in a given time slot, $Prob.(r_n = 0)$ can be expressed as

$$Prob.(r_n = 0) = 1 - \sum_{l \in \{1, 2, 4, \cdots, 2Q\}} Prob.(r_n = l)$$
$$= 1 - \sum_{l \in \{1, 2, 4, \cdots, 2Q\}} \left(\sum_{r=l/2}^Q \sum_{s=0}^{r-l/2} a_{r,s}^{j,1} + \sum_{r=l/2+1}^Q a_{r,1}^{j,2} \right).$$
(5.5)

As we can see from the feasible rate set of TBS, \mathcal{R} , the rate adaptive modulation-assisted TBS scheme does not support odd transmission rate except transmission rate of one, i.e., $r_n \neq 3, 5 \cdots 2Q - 1$. Therefore, transition probabilities for infeasible transmission rates i.e., $r_{l \to k}$ $(l = 3, 5, 7 \cdots, 2Q - 1; \forall k)$ are set equal to zero in Eq. (5.3).

Using the transition probabilities in Eq. (5.3) and assuming an infinite buffer space, the state transition matrix for the DTMC describing the system can be written as Eq. (5.6) in next page.

	г	1		~				1					ı -	1
	·.						•••							
	\mathbf{H}^{I+I}	\mathbf{H}^{0}	$^{-1}$ H	•••	$H^{(2Q-3)-}$	$H^{(5O-5)-}$	$\mathbf{H}^{(5O-1)-}$	\mathbf{H}^{5O-}						
					÷	÷	÷	:						
					\mathbf{H}^{1+}	${ m H}^0$	\mathbf{H}^{I-I}	$-^{z}H$	\mathbf{H}^{3-}		H^{5G-}			
						\mathbf{H}^{I+I}	\mathbf{H}^{0}	$-^{\tau}\mathbf{H}$	H^{5-}	•••	$H^{(5O-1)-}$	$\mathrm{H}^{5 \mathrm{\circ} -}$		
(9.č)						· ·	$\mathbf{H}^{\mathrm{I}+}$	\mathbf{H}^{0}	$H_{(50)}^{1-}$	• • •	${ m H}^{(5{ m C}-5)}_{(5{ m C})}$	${ m H}_{(5 \%)}^{(5 \% - 1)}$	$\mathbf{H}_{(50)}^{50-}$	= d
								$\mathbf{H}^{\mathbf{I}_{+}}$	\mathbf{H}^{0}		${ m H}_{(5O-1)}^{(5O-3)-}$	$\mathbf{H}_{(\mathtt{5G-1})}^{(\mathtt{5G-2})-}$	$\mathbf{H}_{(5O-1)}^{(5O-1)-}$	
										÷	÷	÷	:	
										$\mathbf{H}^{\mathtt{I}+}$	\mathbf{H}^{0}	$\mathbf{H}_{(\mathtt{5})}^{\mathtt{1}_{-}}$	${ m H}_{(5)}^{5-}$	
	,										\mathbf{H}^{1+}	$\mathbf{H}_{(1)}^{0}$	$\mathbf{H}_{(1)}^{1-}$	
	[****						$\mathbf{H}_{(0)}^{\mathbf{I}+}$	${ m H}^{0}_{(0)}$	
	-	I				•		I				·	1	

~)

75

.

By constructing blocks of submatrices as shown in Eq. (5.6), we can obtain a *quasibirth-and-death* (QBD) process transition matrix as follows:

$$\mathbf{P} = \begin{bmatrix} \mathbf{L}_{0} & \mathbf{F}_{0} & & \\ \mathbf{B}_{1} & \mathbf{L}_{1} & \mathbf{F} & & \\ & \mathbf{B} & \mathbf{L} & \mathbf{F} & \\ & & \mathbf{B} & \mathbf{L} & \mathbf{F} & \\ & & & \ddots & \ddots & \ddots & \end{bmatrix}.$$
(5.7)

Now we define a new matrix \mathbf{K}^{j} from matrix \mathbf{M} in Eq. (5.3) by keeping the $(j+1)^{st}$ th row and zeroing out other rows as follows:

$$\mathbf{K}^{(j)} = \begin{bmatrix} 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{r}_{j \to 0} & \mathbf{r}_{j \to 1} & \mathbf{r}_{j \to 2} & \mathbf{r}_{j \to 3} & \cdots & \mathbf{r}_{j \to 2Q} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 \end{bmatrix}$$

With this matrix defined, we can derive the inner submatrices of P in eq. (5.6) as follows:

$$\mathbf{H}_{0}^{(0)} = (1 - \mu)\mathbf{M}$$
(5.8)

$$\mathbf{H}_{1+}^{(0)} = \mu \mathbf{M} \tag{5.9}$$

$$\mathbf{H}_{1^+} = \mu \mathbf{K}^{(0)} \tag{5.10}$$

$$\mathbf{H}_{d^{-}}^{(c)} = \begin{cases} (1-\mu) \sum_{j=c}^{2Q} \mathbf{K}^{(j)}, & \text{if } d = c \\ \mu \sum_{j=c}^{2Q} \mathbf{K}^{(j)} + (1-\mu) \mathbf{K}^{(c-1)}, & \text{if } d = c-1 \end{cases}$$
(5.11)

$$\begin{aligned} \mu_{d^{-}} &= \\ \mu_{j=c} \mathbf{K}^{(j)} + (1-\mu) \mathbf{K}^{(c-1)}, & \text{if } d = c-1 \\ \mu \mathbf{K}^{(d+1)} + (1-\mu) \mathbf{K}^{(d)}, & \text{if } 1 \leq d < c-1 \end{aligned}$$

$$\mathbf{H}_{0}^{(1)} = (1-\mu)\mathbf{K}^{(0)} + \mu \sum_{j=1}^{2Q} \mathbf{K}^{(j)}$$
(5.12)

$$\mathbf{H}_0 = (1-\mu)\mathbf{K}^{(0)} + \mu\mathbf{K}^{(1)}$$
(5.13)

$$\mathbf{H}_{d^{-}} = \mu \mathbf{K}^{(d+1)} + (1-\mu) \mathbf{K}^{(d)}.$$
 (5.14)

5.2.2 Matrix-analytic solution for steady state distributions

 \sim

With the system described by a QBD type chain whose block matrices are derived in the previous subsection, we can now apply the matrix-analytic method [48] to find its steady state distributions. From these steady state probability distributions, we can derive the steady state probability vectors $\pi = [\pi_0, \pi_1, \pi_2, \cdots]$ for queue lengths in the buffer. In fact, the steady-state probability π satisfies the following equations:

$$\pi \mathbf{P} = \pi \tag{5.15}$$

$$\sum_{i=0}^{\infty} \pi_i \mathbf{1} = 1 \tag{5.16}$$

where 1 is a column vector of all ones with appropriate dimension. We can find π_0 , π_1 , and π_2 using the boundary and the normalization conditions. Other values of π_i (i > 2) can be calculated from π_2 through a non-negative matrix **R** as follows: [48]

$$\boldsymbol{\pi}_i = \boldsymbol{\pi}_2 \mathbf{R}^{i-2}.\tag{5.17}$$

Once the steady state vector, π is obtained for the QBD process, we need to partition π_i to obtain the steady-state probability vector of the original transition matrix in Eq. (5.6). Let $N1 = (2Q + 1), N2 = 2Q(2Q + 1), \text{ and } \mathbf{b}_i = [0 \cdots \mathbf{I}_{N1} \cdots 0]^T$ (i = 1, 2, N) be a matrix of size $N2 \times N1$, where \mathbf{I}_{N1} is an identity matrix at the *i*th position. We then user these matrices to obtain the partitions of π . The steady-state probability vector representing the case of having *l* packets in the buffer can be written as:

$$\mathbf{x}_0 = \boldsymbol{\pi}_0 \tag{5.18}$$

$$\mathbf{x}_{l} = \pi_{i}\mathbf{b}_{h}, \quad l \ge 1, \quad (i-1)N < l \le iN \text{ and } h = l - (i-1)N.$$
 (5.19)

5.2.3 Derivation of performance measures

After finding the state vector, we now can determine the performance measures of interest. We are mainly interested about two important performance measures: the queue length distribution and the distribution of queuing delay for packets belonging to the target queue j. We denote by queuing delay the time between the moment a packet arrives at jth buffer and the moment the packet gets access into the medium for transmission. In other words the queuing delay gives a measure of number of times slots that an incoming packet has to wait at the transmission buffer.

Queue length distribution

The marginal probability of finding l packets in the queue can be written as:

$$p_L(l) = \mathbf{b}_l.\mathbf{1} \tag{5.20}$$

where $p_L(l)$ is the probability that the buffer has l packets at an arbitrary time.

Delay distribution

The queuing delay distribution for a packet belonging to the tagged user queue is derived by considering an absorbing Markov chain. The chain initializes from the state when the packet arrives the corresponding queue and gets absorbed when the packet reaches the position where it can be selected for transmission by the subsequent time slot. Note that, in this case, there is no arrival in that particular queue while the process moves towards the absorbing state. Therefore, the transition matrix P_{abs} of the chain can be constructed by setting $\mu = 0$. The next step for finding the delay distribution is to determine the initial probability vector for this absorbing Markov chain. However, since the arrival process is memoryless, the steady state probability vector for *b* packets in the buffer as seen by an arriving packet is π_b . Therefore, the initial probability vector for the Markov chain can be written as

$$\xi^{(0)} = \pi. \tag{5.21}$$

After d time slots the state vector of the absorbing chain is

$$\boldsymbol{\xi}^{(d)} = \boldsymbol{\xi}^{(0)} \mathbf{P}^{d}_{abs} \tag{5.22}$$

where \mathbf{P}_{abs}^{d} is the transition matrix derived by multiplying \mathbf{P}_{abs} with itself *d* times. Assuming that $\boldsymbol{\xi}^{(d)}$ can be partitioned level by level to reach elements $\boldsymbol{\xi}_{x,r}^{(d)}$, we can write the *CDF* of the queuing delay *D*, $P_D(d)$ as

$$P_D(d) = \sum_{\substack{r=0,1,\cdots,2Q\\0 \le x \le r}} \xi_{x,r}^{(d)}.$$
(5.23)

As we mentioned in Chapter 4 that with our proposed HTS scheme finding the probabilities that information packets are transmitted to *j*th user renders joint PDF of the CNRs of the first best user and second user who has the highest normalized CNR except first best user. Since the expression of this joint PDF is not known, we cannot find the rate transition matrix in Eq. (5.3). However, in Section 5.4, we compute buffer distribution and CDF of queuing delay of our proposed HTS scheme via computer simulation.

5.3 Comparison with Classical Single User Scheduling

According to the classical SBS with uncoded adaptive $M_r = 2^{2r}$ -QAM constellations ($r \in \{1, 2, \dots, Q\}$), the scheduler at the BS selects the user (denoted by $U_{(1)}$) who has the largest CNR among all the users. If the CNR of the user $U_{(1)}$, $g_{(1)}$ falls in the *r*th fading region, the constellation size M_r is transmitted. In order to evaluate the buffer distribution and delay statistic, we need to obtain the serving rate transition probabilities in Eq. (5.3). Since the considered SBS scheduling scheme supports even transmission rate with square QAM, the transition probabilities, $r_{l\to k}$ ($l = 1, 3, \dots, 2Q - 1; \forall k$) in Eq. (5.3) are set to zero. For a fading environment where the channel fading gain of a given user varies independently one time slot to the next, the probability $r_{l\to k}$ corresponds to the probability that l packets are transmitted from jth buffer. This probability, $\operatorname{Prob}(r_n = l)$ can be expressed as

$$\operatorname{Prob.}(r_n = l) = a_l^j \tag{5.24}$$

where a_l^j denotes the probability that *j*th user is selected as the first best user and its fading gain falls in the *l*th fading region. The probabilities a_l^j $(l = 0, 2, 4 \dots, 2Q)$ have been evaluated in closed-form in Eq (4.27) in Chapter 4.

According to the NSS (or equivalent PFS) scheme with uncoded adaptive $M_r = 2^{2r}$ -QAM constellations ($r \in \{1, 2, \dots, Q\}$), the scheduler at the BS selects the user who has the largest normalized CNR among all the users. If the instantaneous CNR of the selected user in the *r*th fading region, the constellation size M_r is transmitted. Similarly, we can obtain the transition probabilities in Eq. (5.3). Specifically, the transition probabilities, $r_{l\to k}$ $(l = 1, 3, \dots, 2Q - 1; \forall k)$ in Eq. (5.3) are set to zero and the probability, Prob. $(r_n = l)$ can be expressed as

$$\operatorname{Prob}\left(r_n = l\right) = b_l^j \tag{5.25}$$

where b_l^j denotes the probability that *j*th user has the highest normalized channel fading gain and its instantaneous fading gain falls in the *l*th fading region. The probabilities b_l^j

 $(l = 0, 2, 4 \cdots, 2Q)$ have been evaluated in closed-form in Eq (4.34) in Chapter 4.

Given the fact that the transition probabilities in Eq. (5.3) are obtained for classical SBS and NSS, we can evaluate the buffer distribution and delay distribution of these schemes using the similar methodology that is described in Sections 5.2.2 and 5.2.3.

5.4 Numerical Results

For our numerical results we consider both i.i.d. and i.n.d fading environments. We consider a system of 10 users and maximum constellation size of 1024 which corresponds to maximum transmission rate of 10 packets per slot. We assume that users have a target BER of 10^{-4} . We use MATLAB optimization toolbox to find the rate switching thresholds and statistical toolbox to generate independent Rayleigh fading channel gains for the users. We assume Bernouli packet arrival process to the users' buffers at the BS. We also provide computer simulation results for MMP arrival process.

5.4.1 I.I.D. fading environment

For the i.i.d. fading environment, it is assumed that users have an average channel fading gain of 15[dB] and an identical packet arrival statistic. Since we consider a homogenous system, all users have similar buffer distribution and delay performance. Therefore, we consider user 1 as the tagged user and plot its buffer distribution as well as CDF of queuing delay of packets. For a packet arrival probability of 0.15 for Poisson arrival process, the buffer distributions with our proposed rate adaptive hierarchical modulation-assisted TBS and the classical SBS are plotted in Fig. 5.1 whereas the CDF of delay is plotted in Fig. 5.2. From Fig. 5.1, we observe that our proposed rate adaptive modulation-assisted TBS has higher probability that the buffer will remain at low occupancy region than that of the classical SBS. This will lead to low buffer overflow probability for a memory limited system. Fig. 5.2 clearly shows that the TBS can improve the queuing delay of the packets compared with the classical SBS. In other words, with our TBS scheme the probability that an arriving packet will experience less queuing delay is higher compared than that of the SBS. This is because TBS scheme offers more frequent channel access to the users. Therefore, the probability that an arriving packet waits a longer period of time in the buffer



Figure 5.1: Buffer distribution (i.i.d. fading, Bernouli arrival with packet arrival probability=0.15 and number of users, K=10).

is lower compared with the SBS.

5.4.2 I.N.D. fading environment

For the i.n.d. fading environment, we assume that users are located in 10 different tiers inside the cell. The tiers are uniformly distributed. The average link gain of the far-end user which is located in the cell boundary is assumed to be 20[dB]. We consider that the kth user is located in tier k from the cell boundary and its average CNR, $\bar{g}_k = (10 + \Delta \bar{g} \times k)$ [dB] where $\Delta \bar{g}$ [dB] is the deviation of average link gains of two successive users due to the path-loss. The near-end user is the closest user to the BS and it has the highest average channel fading gain. Since in this i.n.d. fading environment, different users have different average channel fading gains, different users have different buffer distributions and delay performances. Specifically, the near-end user has higher frequency of information access and thus better delay performance. On the other hand, the far-end user experiences low frequency of information access and hence it will have worst delay performance. In our



Figure 5.2: CDF of the queuing delay (i.i.d. fading, Bernouli arrival with packet arrival probability=0.15 and number of users, K=10).



Figure 5.3: Buffer distribution of the near-end user (i.n.d. fading, Bernouli arrival with packet arrival probability=0.4 and number of users, K=10).

comparison, we consider the two extreme scenarios; the far-end and near-end users' queues as the target queues. For the far-end user, the packet arrival probability of Poisson arrival is assumed to be 0.01 whereas for the near-end user this probability is assumed to be 0.4. In Fig. 5.5, we plot buffer distribution for the scheduling schemes under consideration. It is important to mention that the buffer distribution of HTS scheme is simulated for the considered i.n.d. fading environment as we cannot evaluate its performance analytically. From this figure we can easily observe that the HTS scheme offers almost similar buffer distribution as the NSS. On the other hand, rate adaptive modulation-assisted TBS has better buffer distribution than the SBS scheme as buffer distribution is concentrated in the low occupancy region. The CDF of the delay for various schemes are plotted in Fig. 5.6. This figure clearly shows that our proposed HTS significantly improves the delay performance of the far-end user's packet than the TBS and SBS in a i.n.d. fading environment. This is because the HTS scheme selects the second user based on the normalized CNRs of the users and hence the far-end user can get more frequent access to the channel. HTS has certain degradation in delay performance compared with the NSS scheme which significantly



Figure 5.4: CDF of the queuing delay of the near-end user (i.n.d. fading, Bernouli arrival with packet arrival probability=0.4 and number of users, K=10).



Figure 5.5: Buffer distribution of the far-end user (i.n.d fading, Bernouli arrival with packet arrival probability=0.01 and number of users, K=10).

increases the queuing delay for the near-end user as shown in Fig. 5.3. This is because with NSS scheme the far-end user gets more frequent channel access than the HTS at the expenses of reduced frequency of the near-end user.

5.4.3 Simulation results for MMP arrival process

As we mentioned the assumption of Bernouli arrival process simplifies our queuing analysis presented in this chapter. This is because, in a given time slot, only one packet may arrive. However, in what follows with MMP arrival which is an well accepted traffic model for video traffic over wireless network, queuing analysis becomes more complicated as there are may be infinite number of arrivals in a time slot. In addition due to a memory in the packet arrival process, analysis of delay distribution is not straightforward. Queuing analysis for delay distribution for MMP arrival remains an open research problem. However, we obtain buffer and delay distribution of different scheduling schemes with MMP arrival via computer simulation. In our simulation, we use an arbitrary MMP with two



Figure 5.6: CDF of the queuing delay of the far-end user (i.n.d. fading, Bernouli arrival with packet arrival probability=0.01 and number of users, K=10).



. .

Figure 5.7: Buffer distribution (i.i.d. fading, MMP arrival and number of users, K=10).

possible rates $\lambda_1 = 0.1$ and $\lambda_2 = 0.2$ for arrival traffic. An arbitrary transition matrix for the rates, **P**, is assumed to be

$$\mathbf{P}_{\lambda} = \begin{bmatrix} 0.6 & 0.4 \\ 0.2 & 0.8 \end{bmatrix}$$
(5.26)

and the initial distribution of the arrival rate, π_{λ}^{0} is assumed to be

$$\pi_{\lambda}^{0} = [0.35 \ 0.65]. \tag{5.27}$$

For the i.i.d. fading environment, buffer distribution with the considered MMP arrival is plotted in Fig. 5.7 whereas CDF of the delay is plotted in Fig. 5.8. These figures clearly show that our proposed rate adaptive TBS scheduling scheme outperforms classical SBS. This is expected as mentioned in Section 5.4.1

For the i.n.d. fading environment, buffer distribution and delay CDF of the far-end user with the considered MMP arrival are plotted in Figs. 5.9 and 5.10, respectively. From Fig. 5.9, we observe that HTS has quite similar buffer distribution as the NSS and better than the TBS and SBS schemes. Fig. 5.10 clearly shows that HTS has significant improvement



Figure 5.8: CDF of the queuing delay (i.i.d. fading, MMP arrival and number of users, K=10).


Figure 5.9: Buffer distribution of the far-end user (i.n.d. fading, MMP arrival and number of users, K=10).

in delay performance compared to the TBS and SBS schemes as observed with Poisson arrival in 5.4.2. It has certain degradation of delay performance compared with the NSS scheme.

5.5 Conclusion

In this chapter, we have analyzed the buffer distribution and CDF of delay of our earlier proposed rate adaptive modulation-assisted TBS and HTS schemes via queuing analysis. We have also compared these performances with those of the classical single user scheduling schemes. Some selected numerical results showed that for an i.i.d. fading environment for both Bernouli and MMP arrival the TBS outperforms the classical SBS in terms of buffer distribution, average buffer occupancy, delay distribution. In other words, this figure clearly shows that the probability that an incoming packet will be transmitted within d time slots from a given user's buffer is higher than that of the classical SBS. Therefore, TBS can be used in order to improve the radio link layer performances for an i.i.d. fading



Figure 5.10: CDF of the queuing delay of the far-end user (i.n.d. fading, MMP arrival and number of users, K=10).

environment. Selected numerical results in an i.n.d. fading environment have shown that our proposed HTS scheme significantly improve the buffer and delay distribution of the packets at the transmission buffer. This improved delay performance with TBS will help in admitting more number of users than the classical SBS for a given statistical delay constraint. It is important to mention that this improved performance is coming without any degradation of system's overall throughput.



Figure 5.11: Buffer distribution of the near-end user (i.n.d fading, MMP arrival and number of users, K=10).



Figure 5.12: CDF of the queuing delay of far-end user (i.n.d. fading, MMP arrival and number of users, K=10).

Bibliography

- [46] Md. J. Hosssain, M.-S. Alouini, and V. K. Bhargava, "Rate Adaptive Hierarchical Modulation-Assisted Two-user Opportunistic Scheduling" *IEEE Trans. Wireless Commun.*, to appear Jun. 2007.
- [47] L. B. Le, E. Hossain, and A. S. Alfa, "Queueing analysis and admission aontrol for aulti-rate wireless networks with opportunistic scheduling and ARQ-based error control," in *Proc. of IEEE Int. Conf. on Commun. (ICC'05)*, Seoul, Korea, May 2005, pp. 3329 - 3333.
- [48] M. F. Neuts, "Matrix geometric solutions in stochastic models an algorithmic approach", John Hopkins Univ. Press, Baltimore, MD, 1981.

Chapter 6

Hierarchical Constellation for Multi-Resolution Data Transmission over Block-Fading Channels

¹²As we mentioned in Section 1.1.2 that a hierarchical image transmission scheme using a hierarchical OAM constellation was proposed in [50] to protect different resolution components differently. This scheme proposed to devote the first level of hierarchy (i.e., the highest protected sub-channels) of the hierarchical 4/16-QAM constellation for basic information transmission while transmitting refinement information in the second level of hierarchy (i.e., the lowest protected sub-channels). Non-uniform signal constellations in conjunction with ARO scheme were used in [51] to adapt data rate of a single resolution message string transmitting over fading channels. For the same reason as [49], the system proposed in [51] also used a instantaneous CNR threshold-based retransmission scheme. Inspired by this application of non-uniform constellations, we propose, in this chapter, a new multi-resolution data transmission technique over block-fading channels. The proposed scheme uses a threshold-based retransmission strategy to meet a prescribed error performance of the transmitted services while taking advantage of the hierarchical constellation to provide different levels of protection to different resolution levels. In addition, it uses an adaptive scheduling protocol to give different transmission priorities to different resolution levels. Specifically, depending on the scheduling and channel state of the previous transmission, it dynamically selects packets from different resolution levels for the current transmission and the bits from these selected packets are assigned to different hierarchies of a hierarchical 4/16-OAM constellation.

¹²The work presented in this chapter has been published as Md. J. Hossain, M.-S. Alouini, and V. K. Bhargava, "Hierarchical Constellation for Multi-Resolution Data Transmission over Block Fading Channels", *IEEE Trans. Wireless Commun.*, vol. 5, pp.849 - 857, April 2006.

The selection of packets for transmission and the assignment of these selected packets to different hierarchies of the hierarchical constellation are referred to as the scheduling protocol in our proposed scheme. We model this protocol by a finite state first order Markov chain and obtain the packet loss rate and the packet transmission rate over Nakagami-*m* block-fading channel in closed-form. We then we propose a power-distortion optimized real-time multi-resolution video/image data transmission technique which jointly considers the timeliness requirements of the streaming application, the importance levels of different resolution levels and the time-varying channel property. A joint optimization of the proposed transmission framework is proposed by formulating it as a Markov decision process (MDP).

6.1 System and Channel Models

6.1.1 Overall system description

The block diagram of our proposed scheme is shown in Fig. 6.1, where we assume two layer image/video coding, namely, the base information layer (layer 1) and the refinement information layer (layer 2). The level 1 information layer contains the most significant part (MSP) and the information layer 2 contains the least significant part (LSP). According to the scheduling and channel state of the previous transmission, the packet selector selects two packets (namely packet 1 and packet 2 as shown in Fig. 6.1)¹³ to be transmitted for the current transmission. For this transmission, the bits from the selected two packets are assigned to two different hierarchies of the hierarchical 4/16-QAM constellation.

In our proposed scheme, the selection of packets and assignments of these packets to different hierarchies of the hierarchical QAM constellation are performed by the adaptive scheduler as shown Fig. 6.1, whose mode of operation is going to be described in Section 6.2.1.

¹³In general, these two packets may be selected from any resolution (i.e., not necessarily always from resolution 1 and resolution 2, respectively).



Figure 6.1: Block diagram for multi-resolution video/image data transmission.

6.1.2 Hierarchical 4/16-QAM

Our proposed scheme uses hierarchical 4/16-QAM constellation where there are two levels of hierarchies which we refer to as first and second level of hierarchies, respectively. The detailed description of these constellations are given, for example, in [50, 52]. Without going into all these details, a brief review of these constellations is given below. A hierarchical 4/16-QAM constellation, shown in Fig. 6.2, can be modeled as follows. We assume that there are 2 bit streams of data. Each one of these incoming streams carries information of a particular priority. For every channel access two bits are chosen from each level of priority. The two bits from the highest priority stream are assigned to the MSB position in the I and Q channels. Consequently, the two bits from the least priority stream are assigned to the least significant bit LSB position in the I and Q channels. In Fig. 6.2, the fictitious 4-QAM constellation denoted by the black circle represents the highest priority (hereafter referred to as the first level of hierarchy). The distance d_1 represents the first priority. The actual transmitted white symbols represent the 16-QAM constellation. This is the second level of hierarchy and the distance d_2 represents the second priority.

The exact BER expressions over AWGN channel for this hierarchical 4/16-QAM constellation were derived in [52]. For instance, the BER expression for the two bits in the first hierarchy, BER_1 is given by

$$BER_1(g,p) = \frac{1}{2} \left(\frac{1}{2} \operatorname{erfc} \sqrt{\frac{p^2 - 2p + 1}{2(1+p^2)}g} + \frac{1}{2} \operatorname{erfc} \sqrt{\frac{(p^2 + 2p + 1)}{2(1+p^2)}g} \right), \quad (6.1)$$



Figure 6.2: Hierarchical 4/16-QAM constellation.

whereas the BER expression for the two bits in the second hierarchy, BER₂ is given by

$$BER_{2}(g,p) = \frac{1}{2} \left(\operatorname{erfc} \sqrt{\frac{1}{2(1+p^{2})}g} + \frac{1}{2} \operatorname{erfc} \sqrt{\frac{(4p^{2}-4p+1)}{2(1+p^{2})}g} - \frac{1}{2} \operatorname{erfc} \sqrt{\frac{(4p^{2}+4p+1)}{2(1+p^{2})}g} \right), \qquad (6.2)$$

where $p = d_1/d_2$ is the constellation priority parameter, g is the symbol CNR and erfc() is the complementary error function.

From (6.1) and (6.2), it is evident that for a given value of the parameter p, the two hierarchies have two different BER performances for a fixed CNR. Equivalently, the two hierarchies require two different CNRs in order to meet a given target BER or correspondingly a given target PER and these CNR thresholds can be controlled by changing the parameter p.

6.1.3 General assumptions

Our proposed scheme has the following operating assumptions:

A1: We consider a stop-and-wait retransmission strategy i.e., the transmitter sends two packets simultaneously (one in the first level of hierarchy and the other in the second level of hierarchy of the hierarchical constellation) and waits for a reply from the receiver before transmitting the next two packets.

A2: The maximum allowable number of retransmissions of a packet is N_m . Constrained retransmission is employed to meet the delay requirements of the video/image source. In addition, in order to maintain a target PER, we use the same retransmission strategy (i.e., the instantaneous CNR threshold-based strategy) that was used in [49]. However, cyclic redundancy check (CRC)-based ARQ scheme can also be used in conjunction with our proposed scheduling algorithm. In this case, rather than accepting packets below a target PER, data link layer will accept only error-free packets (assuming probability of undetected error is zero) at the expense of redundancy, decoding complexity, and delay. The effects of using CRC-based ARQ scheme will also be briefly discussed in the following sections.

A3: The time duration for simultaneous transmission of two packets is always the same and it is called a time step. For convenience, this discrete time step is denoted by n $(n = 0, 1, \dots)$ in the description of our proposed scheme. We also use the term "transmission" for simultaneous transmission of any two packets during a time step.

A4: The channel fading is assumed to be remain roughly the same over the whole time step. In other words, we are using the so called block-fading channel model as it was used, for example, in [53], [54], [55].

A5: The CNR g is independently distributed from one transmission to another for a specific user. This is, for example, valid for a TDMA multiuser environment because the time duration between two transmissions to a specific user is typically larger than this user's channel coherence time.

6.1.4 Channel model

As we mentioned above, we use a block-fading channel model which is a well accepted model for wireless communication environments with slowly moving terminals. Since the Nakagami-m distribution captures a large class of fading channels, we use it in our block-fading channel model and this implies the following PDF for the received CNR g per block [56]:

$$p_g(g) = \left(\frac{m_n}{g_{0n}}\right)^{m_n} \frac{g^{m_n-1}}{\Gamma(m_n)} \exp\left(-m_n \frac{g}{g_{0n}}\right), \quad g \ge 0$$
(6.3)

where g_{0n} and m_n are the average CNR and the Nakagami fading parameter ($m_n \ge 1/2$) at time step n, respectively and $\Gamma(m) := \int_0^{+\infty} t^{m-1} e^{-t} dt$ is the gamma function.

Since two different hierarchies of the hierarchical 4/16-QAM constellation require two different CNR thresholds to meet a target PER and the instantaneous CNR g is an independent random variable in each time step (according to assumption A5), the channel can be represented by a three state memory less channel. Without loss of generality, let us consider γ_1 and γ_2 are the required CNR thresholds for the first and second level of hierarchy, respectively, with $\gamma_1 < \gamma_2$. The states and the associated probabilities for the three state memory less channel are given below:

State C₁₁: Both packets in the first and second hierarchies are received successfully (i.e., $\gamma_2 \leq g$). The associated probability¹⁴ of receiving both packets successfully p_{11} is

¹⁴The state probabilities are in general function of the time index n. However, since we are interested to evaluate the performance of our proposed scheme for specific average CNR g_0 and fading parameter m, we

thus given by

$$p_{11} = \int_{\gamma_2}^{+\infty} p_g(g) dg = \frac{\Gamma(m, m\frac{\gamma_2}{g_0})}{\Gamma(m)}$$
(6.4)

where $\Gamma(x, a) := \int_x^\infty t^{a-1} e^{-t} dt$ is the complementary incomplete gamma function. **State C**₁₀: The packet transmitted in the first hierarchy is received successfully while the packet in the second hierarchy is not received successfully (i.e., $\gamma_1 \leq g < \gamma_2$). The associated probability p_{10} is thus given by

$$p_{10} = \int_{\gamma_1}^{\gamma_2} p_g(g) dg$$

= $\frac{\Gamma(m, m\frac{\gamma_1}{g_0})}{\Gamma(m)} - \frac{\Gamma(m, m\frac{\gamma_2}{g_0})}{\Gamma(m)}.$ (6.5)

State C₀₀: None of the packets are received successfully (i.e., $\gamma_1 > g$). The associated probability p_{00} is thus expressed as

$$p_{00} = \int_0^{\gamma_1} p_g(g) dg = 1 - \frac{\Gamma(m, m\frac{\gamma_1}{g_0})}{\Gamma(m)}.$$
 (6.6)

When a CRC-based ARQ scheme is used, the channel will have another extra state C_{01} that the packet transmitted in the second hierarchy is received successfully while the packet in the first hierarchy is not received successfully. The probability of this event p_{01} is very unlikely specially with highly asymmetric constellations and consequently we can still use the memory less three states (C_{11} , C_{10} and C_{00}) channel model. In this case, the state probabilities, p_{11}^{CRC} , p_{10}^{CRC} , and p_{00}^{CRC} cannot be expressed in terms of CNR thresholds γ_1 and γ_2 as in (6.4), (6.5) and (6.6) since there is no CNR threshold to meet a target PER. However, they can be expressed in terms of the instantaneous BER expressions as given below:

$$p_{11}^{\text{CRC}} = [1 - \text{BER}_1(g, p)]^{N_p} \times [1 - \text{BER}_2(g, p)]^{N_p}$$
(6.7)

$$p_{10}^{\text{CRC}} = [1 - \text{BER}_1(g, p)]^{N_p} \times [1 - [1 - \text{BER}_2(g, p)]^{N_p}]$$
 (6.8)

$$p_{00}^{\text{CRC}} = [1 - [1 - \text{BER}_1(g, p)]^{N_p}] \times [1 - [1 - \text{BER}_2(g, p)]^{N_p}]$$
(6.9)

where N_p is the packet length in bits per packet and p is the constellation priority parameter as defined before.

drop the time index n without loss of generality.



Figure 6.3: Scheduling state diagram when the maximum number of retransmissions is equal to one (i.e., $N_m = 1$).

6.2 Proposed Protocol

6.2.1 Protocol description

To describe the protocol, let us consider a simple example where we limit the allowed maximum number of retransmissions, N_m to one (i.e., one original transmission plus one retransmission). Initially, we select one packet from each level of information to transmit. Since level 1 information packet requires the highest protection, it is assigned to the first hierarchy of the 4/16-QAM constellation. Correspondingly, the second hierarchy is assigned to level 2 information packet. This is the initial state of our scheduling protocol and as such is denoted by S_{0.0} as shown in Fig. 6.3.

Without loss of generality, let us consider that at the *n*-th time step $(n = 0, 1, \dots)$, the scheduling state is S_{0,0}. Depending on the channel states (C₀₀, C₁₀ and C₁₁), the scheduler has three modes of operation at the (n + 1)-th time step. The three modes of operation are given below.

Channel is in state C_{11} : Both level 1 and level 2 packets are transmitted successfully. In this case, new packet from level 1 in the first hierarchy and a new packet from level 2 in the second hierarchy are scheduled to be transmitted at the (n + 1)-th time step. So, the scheduling for the next transmission remains same as $S_{0,0}$ with a probability of p_{11} as shown in Fig. 6.3. **Channel is in state** C_{10} : Only level 1 packet is transmitted successfully at the *n*-th time step. The level 2 unsuccessfully transmitted packet at the *n*-th time step is discarded and new packet from level 2 in the second hierarchy with a level 1 new packet in the first hierarchy is scheduled to be transmitted at the (n + 1)-th time step. So, there is a loss of level 2 packet and the scheduling remains in the same state $S_{0,0}$ for the next transmission with a probability of p_{10} as shown in Fig. 6.3. It is important to note that the level 2 unsuccessfully received packet is discarded (i.e., no retransmission is requested) because level 1 information packet is given higher priority than level 2 packet.

Channel is in state C_{00} : Both level 1 and level 2 packets fail to be transmitted successfully at the *n*-th time step. In this case, the level 1 old packet is rescheduled for first retransmission in the first hierarchy. Two packets (unsuccessfully transmitted packet at the *n*-th time step and the new packet corresponding to the new level 1 packet selected to be transmitted) are discarded from level 2, and a new packet from level 1 is selected to be transmitted in the second hierarchy. Both packets are selected from level 1 information to give higher transmission priority to level 1 information which is the MSP of the video/image data. This scheduling state is represented by $S_{1,0}$ as shown in Fig. 6.3.

Once the scheduler is in state $S_{1,0}$ at time step (n + 1), again depending on the channel state, three situations may occur. Depending on these three situations, the adaptive scheduler has following mode of operations at the (n + 2)-th time step:

Channel is in state C_{11} : Both level 1 packets are transmitted successfully. In this case, the scheduling for the (n + 2)-th time step is $S_{0,0}$ with a probability p_{11} as shown in Fig. 6.3. **Channel is in state** C_{10} : Only the level 1 packet transmitted in the first hierarchy is received successfully at the (n + 1)-th time step. For the next time step, the level 1 unsuccessfully transmitted packet in the second hierarchy is retransmitted in the first hierarchy to give the highest protection while a new packet from level 1 is transmitted in the second hierarchy. The level 2 packet corresponding to the level 1 new packet just selected for transmission is discarded. So, there is a loss of level 2 packet.

Channel is in state C_{00} : Both level 1 packets fail to be received successfully at the (n+1)th time step. The packet in the first hierarchy is discarded as it has completed the allowed maximum number of retransmission (i.e., one retransmission). So, there is a loss of level 1 packet. The unsuccessfully transmitted packet in the second hierarchy is now assigned to the first hierarchy to give the higher protection and a level 1 new packet is assigned to the second hierarchy to be transmitted in the (n + 2)-th time step. Here again the level 2 packet corresponding to the level 1 new packet just selected for transmission is discarded. So, there is a loss of level 2 packet. After all, the scheduling, for the next step, remains same as $S_{1,0}$ with a probability of p_{00} as shown in Fig. 6.3.

When the maximum number of retransmissions is equal to one, the scheduler has only two states ($S_{0,0}$ and $S_{1,0}$) and it jumps between these two states from one time step to another depending on the channel state as described above. Using the same methodology and assuming that the allowed maximum number of retransmissions is equal to two (i.e., $N_m=2$), we arrive at a scheduling state diagram which is shown in Fig. 6.4. It is important to note that this scheduling state diagram has two extra states, $S_{2,0}$ and $S_{2,1}$, compared with the state diagram shown in Fig. 6.3. The state $S_{2,0}$ shown in this figure represents the second retransmission of a level 1 packet in the first hierarchy and the original transmission of a level 1 packet in the second hierarchy. Similarly, the scheduling state $S_{2,1}$ represents the second retransmission of a level 1 packet in the first hierarchy and first retransmission of a level 1 packet in the second hierarchy. In this case, the scheduler has as such only four scheduling states and makes transitions between these states depending on the channel state.

6.2.2 **Protocol modelling**

From the description of our proposed protocol in the previous section, it is clear that the transmission scheduling at time (n + 1) only depends on the scheduling at time n and the transitions between scheduling states are associated with certain transition probabilities. In addition, the scheduler has a finite number of states for a finite number of allowed maximum retransmissions. In general, for an allowed maximum number of N_m retransmissions, there will be $[N_m(N_m + 1)/2] + 1$ states and the scheduler makes transition between these states from one time step to the next. The protocol, therefore, can be represented by a $[N_m(N_m + 1)/2] + 1$ states first order Markov chain where the states of scheduler represent the states of the Markov chain. All possible states as well as their significance are listed in Table 6.1. For clarity, only the possible transitions out of states $S_{i,j}$ $(i = 1, 2, \dots, N_m - 1 \text{ and } j = 0, 1, \dots, i - 1)$ are shown in Fig. 6.5. The probable

Table 6.1: Significance of the scheduling states

State	Significance
S _{0,0}	Scheduling a level 1 and a level 2 packet in the first and second
	hierarchy, respectively, for original transmission.
$S_{i,j}$	Scheduling a level 1 packet for <i>i</i> -th $(i = 1, 2, \cdots, N_m)$ retrans-
	mission and a level 1 packet for <i>j</i> -th $(j = 0, 1, \dots, i - 1)$ retrans-
	mission in the first hierarchy and second hierarchy, respectively.



Figure 6.4: Scheduling state diagram when the maximum number of retransmissions is equal to two (i.e., $N_m = 2$).



Figure 6.5: Transition out of state $S_{i,j}$ $(0 \le i \le N_m - 1, 1 \le j \le i - 1)$ when the maximum number of retransmissions is equal to N_m .

transitions between all states and the associated probabilities are given below:

$$\begin{split} \mathbf{S}_{i,j}(n) & \xrightarrow{p_{00}} \begin{cases} \mathbf{S}_{i+1,j+1}(n+1), & 1 \le i \le N_m - 1, \ 0 \le j \le i - 1, \\ \mathbf{S}_{j+1,0}(n+1), & i = N_m, \ 0 \le j \le N_m - 1, \end{cases} \\ \mathbf{S}_{i,j}(n) & \xrightarrow{p_{10}} & \mathbf{S}_{j+1,0}(n+1), 1 \le i \le N_m, \ 0 \le j < i - 1, \\ \mathbf{S}_{i,j}(n) & \xrightarrow{p_{11}} & \mathbf{S}_{0,0}(n+1), 1 \le i \le N_m, \ 0 \le j \le i - 1, \end{cases}$$

$$\begin{aligned} \mathbf{S}_{0,0}(n) & \xrightarrow{p_{00}} & \mathbf{S}_{1,0}(n+1), \\ \mathbf{S}_{0,0}(n) & \xrightarrow{p_{10}} & \mathbf{S}_{0,0}(n+1), \\ \mathbf{S}_{0,0}(n) & \xrightarrow{p_{11}} & \mathbf{S}_{0,0}(n+1), \end{aligned}$$

$$\begin{aligned} \mathbf{S}_{0,0}(n) & \xrightarrow{p_{11}} & \mathbf{S}_{0,0}(n+1). \end{cases}$$

where $S_{i,j}(k)$ $(k = 0, 1, 2, \dots)$ is scheduling state at time step k. Using Chapman-Kolomogrov equation, state transition equation can be expressed as

$$\mathbf{P}\,\boldsymbol{\pi}(n+1) = \boldsymbol{\pi}(n),\tag{6.11}$$

where $\pi(k)$ is the $\left[\frac{N_m(N_m+1)}{2}+1\right] \times 1$ scheduling probability distribution vector at time k and **P** is $\left[\frac{N_m(N_m+1)}{2}+1\right] \times \left[\frac{N_m(N_m+1)}{2}+1\right]$ one step transition probability matrix whose elements are obtained from (6.10). For example, the state transition probability matrix, **P**, when the allowed maximum number of retransmissions is equal to two, is given by

$$\mathbf{P} = \begin{bmatrix} p_{11} + p_{10} & p_{11} & p_{11} & p_{11} \\ p_{00} & p_{10} & p_{10} & 0 \\ 0 & 0 & 0 & p_{00} + p_{10} \\ 0 & p_{00} & p_{00} & 0 \end{bmatrix}.$$
 (6.12)

At steady-state condition, (6.11) can be represented by the following matrix equation [[57], Chap. 4]:

$$\mathbf{P}\,\boldsymbol{\pi}^* = \boldsymbol{\pi}^* \tag{6.13}$$

where π^* is the steady-state scheduling probability distribution vector and which is represented as

$$\pi^{*} = \begin{bmatrix} \pi_{0,0}^{*} \\ \pi_{1,0}^{*} \\ \pi_{N_{m},0}^{*} \\ \vdots \\ \pi_{N_{m},N_{m}-1}^{*} \end{bmatrix}_{[\frac{N_{m}(N_{m}+1)}{2}+1] \times 1}$$
(6.14)

where the element $\pi_{i,j}^*$ indicates the probability that the scheduler is at state $S_{i,j}$ at steadystate condition. The steady-state scheduling probabilities, $\pi_{0,0}^*$ and $\pi_{i,j}^*$ $(i = 1, 2, \dots, N_m, j = 0, 1, \dots, i-1)$, can be obtained by solving the set of linear equations in (6.13) and using the fact that summation of all the scheduling probabilities is equal to 1 i.e.,

$$\mathbf{1}\boldsymbol{\pi} = 1 \tag{6.15}$$

where 1 is all one vector of appropriate dimension. Specifically, the first equation in (6.13) gives us

$$(p_{11} + p_{10})\pi_{0,0}^* + \sum_{i=1}^{N_m} \sum_{j=0}^{i-1} \pi_{i,j}^* p_{11} = \pi_{0,0}^*.$$
(6.16)

Substituting (6.4)-(6.6) and (6.15) in (6.16), $\pi_{0,0}^*$ can be solved in closed-form as

$$\pi_{0,0}^* = \frac{\Gamma(m, \frac{m\gamma_2}{g_0})}{\Gamma(m) - \Gamma(m, \frac{m\gamma_1}{g_0}) + \Gamma(m, \frac{m\gamma_2}{g_0})}.$$
(6.17)

It is important to note that the proposed protocol as well as the modelling of the protocol is valid when CRC-based constrained ARQ scheme is used in conjunction with hierarchical constellation except that the closed-form expressions for the steady-state probabilities $\pi_{i,j}^*$ will be rather obtained in terms of p_{11}^{CRC} , p_{10}^{CRC} , and p_{00}^{CRC} .

6.3 Performance Analysis

Based on the Markov chain representation of our proposed protocol and the corresponding steady-state scheduling probabilities, in what follows, we derive the closed-form expressions for packet loss rate, P_{loss} defined as the average number of packets that are discarded per time step at the scheduler and packet transmission rate, P_{tran} defined as the average number of packets successfully transmitted per time step after scheduling step¹⁵

¹⁵In our analysis, we only take into account the packets that have been decided to be transmitted or discarded at the scheduler. In other words, we do not use any packet arrival model and include upper-layer effects. However, the overall packet loss rate can easily be derived by taking into account packet discarding probabilities at upper-layers. In this case, packet transmission rate can be derived from overall packet loss rate.

6.3.1 Average packet loss rate

There is a level 1 packet loss when a level 1 packet is retransmitted for the N_m -th time (i.e., the allowed maximum number of retransmissions) but this packet fails to arrive at the receiver successfully. Thus, the level 1 packet loss rate at steady-state condition, P_{loss}^I can be expressed as

$$\mathbf{P}_{\text{loss}}^{I} = \sum_{j=0}^{N_{m}-1} \pi_{N_{m},j}^{*} \, p_{00} \quad \text{packets/time step},$$
(6.18)

where $\pi_{N_m,j}^*$ is the steady-state probability that a level 1 packet is scheduled for N_m th retransmission in the first hierarchy and a level 1 packet is scheduled for *j*-th ($j = 0, 1, \dots, N_m - 1$) retransmission in the second hierarchy. As mentioned above, this steadystate probability can be obtained by solving the set of linear equations in (6.13) and (6.15). Substituting (6.6) in (6.18), we obtain the following expression for P_{loss}^I as:

$$\mathbf{P}_{\text{loss}}^{I} = \sum_{j=0}^{N_{m}-1} \pi_{N_{m},j}^{*} \frac{\Gamma(m) - \Gamma(m, m\frac{\gamma_{1}}{g_{0}})}{\Gamma(m)} \quad \text{packets/time step.}$$
(6.19)

A level 2 information packet is lost if a level 2 packet is scheduled for transmission but is not successfully transmitted or if a new level 1 packet is assigned to be transmitted in the second hierarchy discarding the corresponding level 2 packet. The joint probability of scheduling a level 2 packet for transmission in the second hierarchy and not receiving the packet successfully is $\pi_{0,0}^*(p_{10} + p_{00})$. The probability of assigning a new level 1 packet in the second hierarchy and discarding the corresponding level 2 packet is given by

$$\mathbf{P}_{\rm dis}^{II} = \pi_{0,0}^* \, p_{00} + \sum_{j=0}^{N_m - 1} \pi_{N_m,j}^* \, p_{00} + \sum_{i=1}^{N_m} \sum_{j=0}^{i-1} \pi_{i,j}^* \, p_{10}. \tag{6.20}$$

Thus, the level 2 packet loss rate, P_{loss}^{II} at steady-state can be expressed as

$$\mathbf{P}_{\text{loss}}^{II} = \pi_{0,0}^* \left(p_{10} + p_{00} \right) + \mathbf{P}_{\text{dis}}^{II} \text{ packets/time step.}$$
(6.21)

Using (6.4)-(6.6), (6.15), (6.17), and (6.20) in (6.21), P_{loss}^{II} can finally be written as

$$P_{\text{loss}}^{II} = \frac{\Gamma(m)[\Gamma(m, m\frac{\gamma_1}{g_0}) + \Gamma(m, m\frac{\gamma_2}{g_0})] - [\Gamma(m, m\frac{\gamma_1}{g_0})]^2 - [\Gamma(m, m\frac{\gamma_2}{g_0})]^2}{\Gamma(m)[\Gamma(m) - \Gamma(m, \frac{m\gamma_1}{g_0}) + \Gamma(m, \frac{m\gamma_2}{g_0})]} + \sum_{j=0}^{N_m - 1} \pi_{N_m, j}^* \frac{\Gamma(m) - \Gamma(m, m\frac{\gamma_1}{g_0})}{\Gamma(m)} \text{ packets/time step.}$$
(6.22)

6.3.2 Average packet transmission rate

The average level 1 packet transmission rate, P_{tran}^{I} at steady-state condition and is given by

$$\mathbf{P}_{\text{tran}}^{I} = p_1 + p_2 + 2p_3 \quad \text{packets/time step}$$
(6.23)

where the probability

$$p_1 = \pi_{0,0}^*(p_{11} + p_{10}) \tag{6.24}$$

is the probability of transmitting one packet from level 1 and this packet is successfully transmitted, the probability

$$p_2 = \sum_{i=1}^{N_m} \sum_{j=0}^{i-1} \pi_{i,j}^* \, p_{10} = (1 - \pi_{0,0}^*) \, p_{10} \tag{6.25}$$

is the probability of transmitting two packets from level 1 and only the packet in the first hierarchy is successfully transmitted, and finally the probability

$$p_3 = \sum_{i=1}^{N_m} \sum_{j=0}^{i-1} \pi_{i,j}^* \, p_{11} = (1 - \pi_{0,0}^*) \, p_{11} \tag{6.26}$$

is the probability of transmitting two packets from level 1 and both of them are successfully transmitted. Using (6.4)-(6.6) and (6.17) in (6.24)-(6.26), the average level 1 packet transmission rate, P_{tran}^{I} , in (6.23) can be expressed, after some manipulations, as

$$\mathbf{P}_{\text{tran}}^{I} = 1 - \frac{1 - 2\Gamma(m, \frac{m\gamma_{1}}{g_{0}}) + [\Gamma(m, \frac{m\gamma_{1}}{g_{0}})]^{2}}{\Gamma(m)[\Gamma(m) - \Gamma(m, \frac{m\gamma_{1}}{g_{0}}) + \Gamma(m, \frac{m\gamma_{2}}{g_{0}})]} \quad \text{packets/time step.} \quad (6.27)$$

The average number of level 2 packet successfully transmitted per time step is the number of packet from level 2 scheduled for transmission per time step multiplied by the probability that the packet is transmitted successfully. Clearly this probability corresponds to p_{11} and as a consequence P_{tran}^{II} can be expressed as

$$\mathbf{P}_{\text{tran}}^{II} = \pi_{0,0}^* p_{11} \qquad \text{packets/time step.}$$
(6.28)

Using (6.4) and (6.17) in (6.28), P_{tran}^{II} can be written as

$$P_{\text{tran}}^{II} = \frac{\left[\Gamma(m, \frac{m\gamma_2}{g_0})\right]^2}{\Gamma(m)[\Gamma(m) - \Gamma(m, \frac{m\gamma_1}{g_0}) + \Gamma(m, \frac{m\gamma_2}{g_0})]} \qquad \text{packets/time step.}$$
(6.29)

When CRC-based ARQ strategy is used in our proposed scheduling algorithm, it is straightforward to obtain the closed-form expressions for packet transmission rate and packet loss rate for both resolutions in terms of p_{11}^{CRC} , p_{10}^{CRC} , and p_{00}^{CRC} which are functions of instantaneous CNR g. Therefore, packet loss rate and packet transmission rate will be function of the instantaneous CNR g and their average value can be obtained simply by taking the expectation with respect to the CNR g.

6.4 Numerical Results

In this section, we present some numerical results to show the performance of our proposed scheme. Specifically, we plot packet loss rate and packet transmission rate as a function of average CNR g_0 for different values of Nakagami fading parameter m, different constellation priority parameter p, and different allowed maximum number of retransmissions N_m . In our numerical results, we used an uncoded average packet error rate, $PER_0=10^{-3}$ and a packet length of 1000 bits (i.e., $N_p = 1000$ bits/packet). So, the target uncoded average BER, $BER_0 = 1 - (1 - PER_0)^{1/N_p} = 10^{-6}$. The CNR thresholds, γ_1 and γ_2 are found by solving (6.1) and (6.2), respectively, numerically for a given priority parameter p. We used MATLAB Statistical Toolbox to get simulation results which we compared with our analysis-based numerical results.

6.4.1 Average packet loss rate

The average packet loss rate over Rayleigh fading channel (m = 1) is plotted in Fig. 6.6 for different constellation priority parameter p (or equivalently different CNR thresholds). We limit the allowed maximum number of retransmissions, N_m to two. This figure shows that as the value of p increases, the packet loss rate for level 1 information decreases while that of level 2 information increases. This is expected as the value of p increases, the difference between the two CNR thresholds, γ_1 and γ_2 , increases and consequently the packet in the first level of hierarchy gets higher priority than the packet in the second hierarchy. Level 1 information, therefore, gets higher transmission priority than the level 2 information.

The average packet loss rate over Rayleigh fading channel is plotted in Fig. 6.7 for different allowed maximum number of retransmissions. This figure shows that as the allowed



Figure 6.6: Effect of the constellation priority parameter p on the average packet loss rate $(N_m=2 \text{ and } m=1)$.

maximum number of transmissions, N_m increases, the level 1 packet loss rate decreases. This is also expected since a level 1 information packet gets more time to be transmitted successfully before it is discarded as the number of allowed maximum transmission increases.

The average packet loss rate, for different values of Nakagami fading parameter m with a given priority parameter p and allowed maximum two retransmissions, is shown in Fig. 6.8. From this figure, it is evident that at high average CNR regions as the value of mincreases, the packet loss rate for both level 1 and level 2 information decreases. This is due to the fact that the amount of fading decreases as the value of m increases. On the other hand, at low average CNR regions the packet loss rate for both level information packets increases as the value of m increases. This is because the shape of the Nakagami mdistribution becomes narrower centred around the average CNR as the value of m increases [56]. Therefore, at low average CNR and at higher value of m, the probability having of



Figure 6.7: Effect of the allowed maximum number of retransmissions, N_m on the average packet loss rate ($m = 1, p = 5, \gamma_1 = 15.39$ [dB], and $\gamma_2 = 27.69$ [dB]).



Figure 6.8: Effect of the Nakgami m fading parameter on the average packet loss rate $(N_m=2, p=5, \gamma_1=15.39[dB], \text{ and } \gamma_2=27.69[dB]).$



Figure 6.9: Effect of the constellation priority parameter p on the average packet transmission rate ($N_m=2$ and m=1).

better channel condition is lower and consequently packet loss rate increases.

6.4.2 Average packet transmission rate

The average packet transmission rate over Rayleigh fading channel (m = 1) and for different constellation priority parameter, p is plotted in Fig. 6.9 for $N_m = 2$. This figure shows that as the value of p increases, the packet transmission rate for level 1 information increases while that of level 2 information decreases. Again, this is expected since the level 1 information gets a higher transmission priority than the level 2 information as the priority parameter p increases.

Finally, the average packet transmission rate plotted in Fig. 6.10 for a given priority parameter, p, shows that at high average CNR regions as the value of m increases, the packet transmission rate increases. At low CNR regions, we got different scenario; as the value of m increases, the packet transmission rate decreases. These are again expected since the amount of fading decreases as the value of m increases and at low average CNR region, the



Figure 6.10: Effect of the Nakgami *m* fading parameter on the average packet transmission rate (N_m =2, p = 5, $\gamma_1 = 15.39$ [dB], and $\gamma_2 = 27.69$ [dB]).

probability of having better channel quality decreases as the value of m increases.

6.5 Real-time Multi-Resolution Data Transmission over Correlated Fading Channels

¹⁶In Section 6.2, we have proposed an adaptive scheduling protocol for multi-resolution image/video data transmission over a time-varying link. This protocol used a fixed transmit power in each time slot and a retransmission scheme with a maximum allowed number of retransmissions. This retransmissions limit is assumed to be dictated by the delay requirement of the packets that arrive at the transmission buffer. In other words, it did not consider any explicit delay requirements for the arrival packets. In addition, we have also considered that the fading gain varies independently from one time slot to the next.

Intuitively, using higher number of retransmissions for more important packets should provide a better video quality. However, the allowed number of retransmissions for the current packet affects the residual communication resources for the subsequent packets in a streaming application where the packets have a real-time deadline constraint. Specifically, if the sender chooses a higher number of chances for the current packet, the chances of successfully receiving that packet increases. However, the packets that are waiting in the buffer to be transmitted are left with less chances before the presentation deadline at the receiver. This may cause the sender to use smaller number of retransmissions or to discard some packets without being sent as their deadlines have been expired. Therefore, the choice of the number of retransmissions or transmission opportunities for individual packets are not independent. The problem is even more complex in a multi-resolution data transmission system because the system has to consider not only the individual deadline of the packets waiting at the buffer but also their importance levels. For example, let us consider two packets are waiting at the buffers. One is from higher importance level but has less closer deadline and the other is from lower importance level but has more closer deadline. If the transmitter can transmit only one packet per transmission block which packet should

¹⁶This Section has been published as Md. J. Hossain, M.-S. Alouini, and V. K. Bhargava, "Real-time Multi-Resolution Data Transmission over Correlated Fading Channels using Hierarchical Constellations', in *Proc. of IEEE Veh. Technol. Conf. (VTC'06 Spring)*, vol. 5, pp.2068-2072, May 2006.

be selected by the packet scheduler to be transmitted for current transmission in order to achieve a given quality of the received image/video? In addition, the transmitter needs to select the transmit power level according to the time-varying property of the wireless channel, the importance level of the packets and their deadline if there is a given power budget. On the other hand, if the transmitter wants to transmit both packets simultaneously using a hierarchical constellation, which is capable of providing different level of protection to the transmitted messages [50], [52], how much relative protection levels should be given to the individual packets at the physical layer.

Motivated by the aforementioned challenging tasks, in this section, we propose a powerdistortion optimized real-time multi-resolution video/image data transmission technique which jointly considers the timeliness requirements of the streaming application, the importance levels of different resolution levels and the time-varying channel property. Basically, depending on the current buffer occupancy and channel state, it dynamically selects transmit power level, a constellation priority parameter as well as packets from different resolution levels to be transmitted for the current time transmission block. A joint optimization of the proposed transmission framework is proposed by formulating it as a MDP. The power-distortion trade-offs curve is plotted for a given fading rate. We also compare the proposed scheme with a scheme employing uniform signal constellation.

In a related work, a multi-layer video transmission technique using non-uniform phase shift keying (PSK) constellation was proposed in [58]. The scheme in [58] proposed to select the constellation parameter as well as the coding rate according to the wireless channel quality. However, it did not consider any buffering of the packets with a real-time deadline constraint and the correlation among the fading gains from one time slot to the next.

6.5.1 Overall system description

The block diagram of our proposed scheme is shown in Fig. 6.11, where we assume two layer image/video coding, namely, the base information layer, (denoted by R_1) and the refinement information layer (denoted by R_2). The R_1 information layer contains MSP and the R_2 information layer contains the LSP. We consider that packets from both resolutions arrive in a batch i.e., every time slot two packets arrive simultaneously. The R_1 information packet waits at buffer 1 and the corresponding R_2 information packet waits at buffer 2 at



Figure 6.11: Adaptive system for layered real-time image/video data transmission.

the transmitter. For a real-time deadline constraint, packets can wait only two time-slot after they arrive at the buffer¹⁷. More specifically, the packet that arrives in a time-slot n can be transmitted only in the time-slot n or (n + 1) other wise it is discarded due to missing presentation deadline at the receiver. According to channel state and buffer occupancy of both resolutions, the packet selector selects two packets (namely packet 1 and packet 2, as shown in Fig. 6.11). In general, these two packets may be selected from any resolution, not necessarily always from R₁ and R₂ information, respectively. For this transmission, the bits from the selected two packets are assigned to two different hierarchies (namely, H₁ and H₂) of the hierarchical 4/16-QAM constellation. The selection of packets for transmission and the assignment of these selected packets to different hierarchies of the hierarchical constellation are referred to as the packet scheduling protocol in the proposed transmission system. Finally, the transmit power level, P and the constellation priority parameter, p are selected and accordingly the symbols are transmitted from hierarchical 4/16-QAM modulator in a given time-slot.

Wireless channel model

In a time-varying wireless channel, the quality of the channel at a given time is correlated to the previous channel conditions and this kind of behavior can be modelled as a finite state Markov chain (FSMC) [59]. FSMC models can be constructed by partitioning the

¹⁷For simplicity, we consider a two-time slot deadline. However, more than two-time slot deadline can be used in the proposed transmission framework and in that case the complexity of the system increases.

fading gain value into a finite set of ranges. Let us consider that the fading power gains are partitioned into a set of Q ranges where $\mathcal{H} = \{c_1, c_2, \cdots, c_Q\}$ denote the Q states of the FSMC and $\bar{\phi} = [\phi_1, \phi_2, \cdots, \phi_{Q+1}]^T$ are the fading gain thresholds in increasing order with $\phi_1 = 0$ and $\phi_{Q+1} = \infty$. Then the fading gain ranges $[\phi_l, \phi_{l+1}), l = 1, 2, \cdots, Q$ are mapped into the channel states c_l , $l = 1, 2, \dots, Q$, respectively. The channel fading gain corresponding to state c_l is denoted by h_l and obtained by averaging the fading gain over the range $[\phi_l, \phi_{l+1})$. Several methodologies can be used for partitioning the fading gain (e.g., see [59]) and calculating the state transition probabilities. We use the equal-probability method i.e., the fading gain of the channel is partitioned such as the stationary probability associated with all the states is equal [59]. The channel fading gain is assumed to remain roughly the same over the whole transmission block. In other words, we are using the so called block-fading channel model as it was used, for example, in [53], [55]. We assume that the channel states associated with consecutive blocks are assumed to be neighboring states. Therefore, the channel Markov process can be described as a birth-death process. For a given transmission rate and maximum doppler frequency f_d these probabilities can be found in [59]. Since two packets are transmitted in two different hierarchies of the hierarchical 4/16-QAM constellation in a transmission block, the transmitter can have four different observation signals. These observations as well as their associated probabilities are given below.

Observation O^1 : Both packets in hierarchies H₁ and H₂ are received successfully.

Observation O^2 : The packet transmitted in hierarchy H_1 is received successfully while the packet in hierarchy H_2 is not received successfully.

Observation O^3 : The packet transmitted in hierarchy H_2 is received successfully while the packet in hierarchy H_1 is not received successfully.

Observation O^4 : None of the packets is received successfully.

Corresponding probabilities of these observations for a transmit power level P^i , a constellation priority parameter p^j and, a fading power gain h_i of the channel are thus given

$$Prob.(O^{1}|\mathbf{P}^{i}, p^{j}, h_{i}) = [1 - BER_{1}(\mathbf{P}^{i}, p^{j}, h_{i})]^{N_{p}} \times [1 - BER_{2}(\mathbf{P}^{i}, p^{j}, h_{i})]^{N_{p}}, (6.30)$$

$$Prob.(O^{2}|\mathbf{P}^{i}, p^{j}, h_{i}) = [1 - BER_{1}(\mathbf{P}^{i}, p^{j}, h_{i})]^{N_{p}}$$

$$\times [1 - [1 - BER_{2}(\mathbf{P}^{i}, p^{j}, h_{i})]^{N_{p}}] \qquad (6.31)$$

Prob.
$$(O^3 | \mathbf{P}^i, p^j, h_i) = [1 - [1 - \text{BER}_1(\mathbf{P}^i, p^j, h_i)]^{N_p}] \times [1 - \text{BER}_2(\mathbf{P}^i, p^j, h_i)]^{N_p},$$
 (6.32)

Prob.
$$(O^4 | \mathbf{P}^i, p^j, h_i) = [1 - [1 - \text{BER}_1(\mathbf{P}^i, p^j, h_i)]^{N_p}] \times [1 - [1 - \text{BER}_2(\mathbf{P}^i, p^j, h_i)]^{N_p}]$$

(6.33)

where N_p is the packet length in bits per packet, BER₁(P^{*i*}, p^{*j*}, h_{*i*}) and BER₂(P^{*i*}, p^{*j*}, h_{*i*}) are the BERs of the bits in the first and second hierarchies, respectively (given in Eqns. (6.1) and (6.2), respectively).

Markov decision processes

A MDP is a model for deciding how to act in a stochastic environment where outcomes are uncertain. A MDP is described through a set of states $S = \{s^1, s^2, \dots, s^N\}$, a set of actions $\mathcal{A} = \{a^1, a^2, \dots, a^U\}$, a set of state and action dependent immediate costs $c: \mathcal{K} \mapsto \mathbb{R}$, and a set of state and action dependent transition probabilities. Set $\mathcal{K} = \{(s, a)\} : s \in S$, $a \in \mathcal{A}_s$ is the set of state-action pairs. While the system is in state $s_n = s^i \in S$ at decision epoch or time-slot n, the controller or decision maker selects an action a_n from the set of actions \mathcal{A}_{s^i} available at state $s^i \in S$, where $\bigcup_{s^i \in S} \mathcal{A}_{s^i} = \mathcal{A}$. In the next time-slot, the system moves to a state $s^j \in S$ according to the transition probability $p(s_{n+1} = s^j | s_n = s^i, a_n)$ and the decision maker incurs a one step immediate cost $c(s_n, a_n)$.

A policy as a mapping from state to actions specifies the decision rule to be used at all time slot or decision epoch. With a policy φ , which does not depend on time slot, n (called as stationary policy), the expected long-term average cost per stage is

$$J_{\varphi} = \lim_{M \to \infty} \frac{1}{M} \sum_{n=1}^{M} \mathsf{E}_{s}[c(s_{n}, \varphi(s_{n}))], \tag{6.34}$$

where expectation, E_s is over the random evolution of the states of the MDP. An optimal policy denoted by φ^* is the one that incurs the minimum average cost per stage i.e.,

$$\varphi^* = \arg\min_{\varphi \in \Pi} J_{\varphi} \tag{6.35}$$

by



Figure 6.12: Buffer state space.

where Π is a set of all stationary admissible deterministic policies. In general, this minimization should be performed over the set of all causal policies. However, for our transmission frameworks, optimal policy for the resulted MDP exists and is contained within the set of stationary deterministic policies [60]. The optimal policy for an MDP can be found in different ways for example using the relative value iteration algorithm (RVI) or the policy iteration (PI) algorithm [61].

6.5.2 **Problem formulation**

Buffer dynamics, packet scheduling, and composite state space

Since a packet can wait at most two time-slot at the buffer, each buffer can contain at most two packets. To describe the dynamics of the buffer, without loss of generality, let us consider that, at the *n*-th time slot a packet from each resolution, is waiting at its respective buffer as shown in Fig. 6.12. The number of time slots they can wait at the buffer before the presentation deadline is two (i.e., n or (n + 1)-th time slot). As such this buffer state is denoted by B¹ and the notation $R_i^j(n)$, in general, denotes that the packet is from *i*-th resolution level of *n*-th batch and can wait at most *j* time-slot at the buffer. At this state of buffer occupancy since level 1 information packet requires the highest protection, it is assigned to hierarchy H_1 of the hierarchical 4/16-QAM constellation. Correspondingly, hierarchy H_2 is assigned to the level 2 information packet. This packet scheduling option is represented by a circle in Fig. 6.12. Depending on the acknowledgement signal, O_n at time-slot *n*, buffers can have the following states of occupancy in the next time step:

 $O_n=O^1$: Both level 1 and level 2 packets of *m*-th batch are transmitted successfully. A new batch (i.e., (n + 1)-th batch) arrives at the buffers and the buffers remain at the same state B¹ as shown in Fig. 6.12.

 $O_n=O^2$: Only level 1 packet, which was assigned to hierarchy H₁, is transmitted successfully. Therefore, the corresponding level 2 packet of *n*-th batch remains in the buffer. As such this packet is denoted as $R_2^{1+}(n)$ in Fig. 6.12. Here the "+" sign is to denote that the corresponding (level 1 or level 2) packet has been received successfully at the receiver. In addition, the new batch of packets arrives at the buffer. This buffer state is denoted as B^2 .

 $O_n = O^3$: Only the level 2 packet of *m*-th batch is transmitted successfully and a new batch arrives at the buffer. The level 1 packet from *n*-th batch, waiting at the buffer, is denoted as R_1^{1+} . This buffer state is represented by B^3 as shown in Fig. 6.12.

 $O_n = O^4$: None of the packets of *n*-th batch are transmitted successfully and a new batch arrives at the buffer. As such this buffer state is denoted by B⁴ in Fig. 6.12.

Once the buffer is in state B¹ at time-slot (n + 1), the packet selection procedure is straight forward as mentioned above. However, if the buffers are in state B², B³, or B⁴, the buffers have more than two packets to be transmitted and these packets have different importance levels as well as different deadlines. Since the transmitter can send only two packets per time-slot, it has a number of packet scheduling options in these buffer states. For example, at state B², the R₂ information packet of the *n*-th batch is waiting at the buffer which has less priority in terms of received quality of the image or video as already the corresponding R₁ information packet is received correctly. If the packet R¹⁺₁(*n*) is selected to be transmitted, one of the packets from (n + 1)-th batch cannot be transmitted and consequently the number of transmission opportunities before the presentation deadline for this packet reduces to one. In general, the transmitter has $\binom{k}{2}$ packet selection options when there are total k packets in the buffers and two of them are selected to be transmitted. In addition, selected two packets can be assigned to two different hierarchies (H₁ and H₂) of the hierarchical constellations. Therefore, there are $\binom{2\times k}{2}$ scheduling options when the buffer has k packets. For notational convenience, let us consider that the scheduling options that are available at buffer states B^i , i = 1, 2, 3, 4 is denoted by a set \mathcal{D}_{B^i} , i = 1, 2, 3, 4, respectively. Using the same methodology as described above (also in Fig. 6.12), it can be observed that with any available packet scheduling option, the buffers can have only these four states and switches between one to another depending on the acknowledgement signals. The transition probabilities of the buffer states depend on transmit power level, constellation priority parameter, channel state, and the packet scheduling option. Although it is not possible to represent the buffer state transition probabilities in generic form, they can be found by constructing the tree diagram for each buffer state as well as packet scheduling option in this state as shown in Fig. 6.12. To represent the transmission framework in terms of an MDP model, the state space of this model is $S = \mathcal{B} \times \mathcal{H}$ where $\mathcal{B} = \{B^1 \ B^2 \ B^3 \ B^4\}$ is the state space of the buffers' occupancy. Note that the total number of states in S, N is equal to $4 \times Q$.

Action and state pairs

We assume that the transmission power level $P_n = P^j$ at time-slot n is chosen among I discrete power levels of the set $\mathcal{P} = \{P^1, P^2, \dots, P^I\}$ and that the constellation priority parameter $p_n = p^i$ is selected from among J priority parameters of the set $\mathcal{L} = \{p^1, p^2, \dots, p^J\}$. Then the action space for state s^i , $\mathcal{A}_{s^i} = \mathcal{D}_{\phi(s^i)} \cup \mathcal{P} \cup \mathcal{L}$ where the function $\phi(s)$ returns the buffer state of the composite state s.

6.5.3 Power-distortion model

Since our objective is to find an optimal policy which minimizes the transmission power so that the average distortion in the received application remains below certain bounds, the immediate objective cost function of the associated MDP can be given as only the power cost i.e., the transmitted power level in each state. We use a very generalized distortion model. Similar type of distortion model was used in [62]. In general, the distortion for the packets of n-th batch is given by

$$\epsilon = \begin{cases} 0 & \text{if both packets are received successfully,} \\ \Delta c^1 & \text{if only } R_1(n) \text{ packet is received successfully,} \\ \Delta c^2 & \text{if only } R_2(n) \text{ packet is received successfully,} \\ \Delta c^3 & \text{if none of the packets are received successfully.} \end{cases}$$
(6.36)

Therefore, the expected distortion can be expressed as

$$\mathbf{D}^{\text{avg}} = p_{10}\Delta c^1 + p_{01}\Delta c^2 + p_{00}\Delta c^3, \tag{6.37}$$

where p_{10} is the probability of receiving R₁ packet of *i*-th batch only, p_{01} is the probability of receiving R₂ packet of *i*-th batch only and p_{00} is the probability of receiving none of the packets of *i*-th batch. Since R₂ packet of a given batch is not useful without the corresponding R₁, a high value is assigned to Δc^2 .

6.5.4 Results

We formulate the problem as an unconstrained Markov decision process (UMDP) with a trade-off factor. The immediate cost function under a stationary deterministic policy φ is

$$c(s_n, \varphi(s_n)) = \mathbf{P}_n + \beta \mathbf{D}_n^{\text{avg}}$$
(6.38)

where β (> 0) is a trade-off factor and by changing its value we can obtain different trade-offs between average transmit power and the average distortion. For a given β , the stationary deterministic optimal policy φ^* is obtained using relative value iteration. In Fig. 6.13, we plot the power-distortion trade-offs curves of optimal transmitter for a given fading rate (i.e., doppler frequency f_d) using the following settings: average channel gain $h_0 =$ 10[dB], and the noise power $\omega^2 = 1$ [mW]. Two power levels, namely $P^1 = 2$ [mW] and $P^2 = 20$ [mW] and three priority parameter options, namely $p^1 = 2$ (uniform constellation), $p^2 = 3$, and $p^3 = 4$ are used in our simulations. In the same figure, we have also plotted power-distortion trade-offs using a uniform constellation-based scheduling scheme. With the uniform constellation-based scheme a uniform 16-QAM constellation is used i.e., the priority parameter p = 2. The figure shows that as the distortion cost increases, the required average transmit power reduces. This is expected as we increase the transmit power level, the chances of receiving the packets successfully increases. It is also evident from Fig. 6.13 that the uniform constellation-based scheduling scheme requires higher power than the hierarchical constellation-based scheme for the same distortion cost. This is due to the fact that hierarchical constellation can offer different level of protections to the transmitted packets according to their priorities at the physical layer by changing the parameter p.



Figure 6.13: Comparison of hierarchical and uniform constellation-based schemes.

6.6 Conclusion

In this chapter, a multi-resolution transmission system using hierarchical constellation in conjunction with a CNR threshold-based retransmission strategy has been proposed. The proposed scheme not only takes advantage of hierarchical constellation but also uses an adaptive scheduling protocol to give different transmission priorities to different resolution levels. A Markov chain model of this protocol leads to closed-form expressions for the average packet loss rate and the average transmission rate. Numerical results, which were validated by computer simulations, show that the proposed scheme can control the relative transmission priorities in terms of transmission rate as well as loss rate of different resolution levels by varying the asymmetry of the constellation and the maximum number of allowed retransmissions. Then we have proposed an integrated packet scheduling and unequal error protection strategy for real-time packetized multi-resolution (layered) image/video data transmission over correlated channel. In order to select transmit power level, a constellation priority parameter and the packets from different resolution levels of
different packets and time-varying channel property. A joint optimization of the proposed transmission framework has been done formulating it as a MDP. The proposed hierarchical scheme has also been compared with an optimized scheme which uses uniform signal constellation. The scheme with uniform constellation requires higher average transmit power than the hierarchical scheme in order to maintain the same average distortion.

Bibliography

- [49] Q. Liu, S. Zhou, and B. Giannakis, "Jointly adaptive modulation and packet retransmission over block-fading channels with robustness to feedback latency," in *Proc. Conf. Information Sciences and System (CISS 2003)*, The Johns Hopkins University, USA, Mar. 2003.
- [50] M. Morimoto, M. Okada, and S. Komaki, "A hierarchical image transmission system in fading channel," in *Proc. IEEE Int. Conf. Univ. Personal Commun. (ICUPC' 95)*, Tokyo, Japan, Oct. 1995, pp. 769-772.
- [51] R. Otnes and T. Maseng, "Adaptive data rate using ARQ and non-uniform constellations," in *Proc. IEEE Veh. Technol. Conf. (VTC' 01-Spring)*, Rhodes, Greece, May 2001, pp. 1211-1215.
- [52] P. K. Vitthaladevuni and M.-S. Alouini, "BER computations of 4/M-QAM hierarchical constellations," *IEEE Trans. Broadcasting*, vol. 47, no. 3, pp. 228-239, Sep. 2001. See also correction in *IEEE Trans. Broadcasting*, vol. 49, pp. 408, Dec. 2003.
- [53] E. Biglieri, G. Caire, and G. Taricco, "Limiting performance of block-fading channels with multiple antennas," *IEEE Trans. Inform. Theory*, vol. 47, no. 4, pp. 1273-1289, May 2001.
- [54] E. Malkamäki and H. Leib, "Coded diversity schemes on block-fading Rayleigh channels," in *Proc. IEEE Int. Conf. Univ. Personal Comm. (ICUPC' 97)*, San Diego, CA, Oct. 1997 pp. 494-499.
- [55] L. H. Ozarow, S. Shamai, and A. D. Wyner, "Information theoretic considerations for cellular mobile radio," *IEEE Trans. Veh. Technol.*, vol. 43, pp. 359-378, May 1994.

- [56] M. Nakagami, "The *m*-distribution-A general formula of intensity distribution of rapid fading," in *Statistical Methods in Radio Wave Propagation.*, Oxford, UK: Pergamon, 1960, pp. 3-36.
- [57] F. Gebali, *Computer Communication Networks: Analysis and Design*, NorthStar Digital Design, Inc., Victoria, BC, Canada, 2002.
- [58] Y. Pei and J. W. Modestino, "Multi-layered video transmission over wireless channels using an adaptive modulation and coding scheme," in *Proc. Int. Conf. Image Processing (ICIP'01)*, Thessaloniki, Greece, Oct. 2001, pp. 1009-1012.
- [59] H. S. Wang and N. Moayeri, "Finite-state Markov channel-A useful model for radio communication channels," *IEEE Trans. Veh. Technol.*, vol.44, pp. 163-171, Feb. 1995.
- [60] M. L. Putterman, "Markov decision processes : discrete stochastic dynamic programming," New York : John Wiley & Sons, 1994.
- [61] D. P. Bertsekas, *Dynamic programming and Optimal Control*. 2nd ed. vol. I and II, Belmont, Massachusetts: Athena Scientific, 2001.
- [62] L. Yan and N. Bambos, "Power-controlled wireless links for video streaming applications," in *Proc. IEEE Veh. Technol. Conf. (VTC'04 Fall)*, Los Angeles, CA, Vol. 4, Sep. 2004, pp. 2577-81.

Chapter 7

Link Adaptation for OFDM Systems for Delay-Constrained Communication over Correlated Fading Channels

¹⁸As we have mentioned in Chapter 1 that in order to take advantage of time-varying nature of subcarrier fading gains in an OFDM system, link adaptation has been proposed in wireless communications standard. There are two extremes of bit-loading algorithms from the perspective of delay-constrained communication over fading channels. One is hard delay-constrained case and the other is no delay constrained case. However, the arrived data packets may tolerate some intermediate queuing delay at the transmission buffer rather than having hard delay constraint or no delay constraint. Recently channel and buffer adaptive transmission techniques, which schedule packet transmission in order to meet a target delay constraint while minimizing average transmit power, have been proposed for single carrier case [64], [65], [66]. Transmission rate has been optimized via a cross layer optimization technique and the optimal trade-offs between different conflicting objectives has been studied formulating the problem as a MDP. Best to our knowledge, despite that the performance of buffer adaptive multi-carrier transmission scheme has been analyzed in [70], no study has been done on multicarrier loading algorithms that minimize the aggregate BER with a fixed transmit power or minimize the total average transmit power with a target error probability for a given delay-constrained communication over correlated fading channels.

In this chapter, we study and offer some novel results on adaptive loading algorithms for OFDM systems for delay-constrained services. Specifically, we propose and study two

¹⁸The research presented in this chapter is to be revised and submitted as Md. J. Hossain, D.-V. Djonin, and V. K. Bhargava, "Delay limited power and bit-loading algorithms for OFDM systems for delay-constrained communications", *IEEE Trans. Veh. Technol.*

different transmission frameworks for an OFDM system. In the first transmission framework, a bit-loading algorithm, which minimizes the aggregate BER of an OFDM system while maintaining a target buffering delay of the packets in the transmission buffer is proposed. In this framework, we assume a fixed total transmit power which is equally divided among the OFDM subcarriers. In the second transmission framework, we propose bit and power loading scheme with the objective to minimize the average total transmit power while maintaining a target buffering delay of the packets. Due to the system dynamics both loading frameworks are formulated as MDP and the optimal loading policies are obtained via equivalent linear programming (LP) methodology. The complexity of algorithms used to find the optimal loading policies and their implementation issues are described. Since finding optimal policies is complex and practically un-realizable for large number of carriers in the system, we propose a sub-optimal loading algorithm using the results of the single carrier system's optimal loading policy and a greedy approach. Finally, in this chapter, we study the effect of correlation among the fading gains of subcarriers of a multicarrier modulation (MCM) system.

Notation: Subscripts are reserved to denote a specific state of a state space, while superscripts are reserved to denote the subcarrier index. Subscript in brackets denotes the state at a certain transmission slot.

7.1 System Model

A system model for buffer and channel adaptive transmission system is shown in Fig. 7.1 where we assume a single class of traffic is to be transmitted over fading channels using an OFDM system with L subcarriers. It is assumed that each subcarrier undergoes frequency flat fading. Data packets, which arrive from higher application layers, are stored in a finite size transmission buffer of length B in order to be transmitted later. In each transmission slot, the transmitter adaptively assigns a number of bits to different subcarriers jointly considering the state of buffer occupancy as well as different subcarriers' states. In what follows we provide detailed descriptions of the channel model for OFDM subcarriers as well as incoming traffic and rate (or both rate and power) adaptation models which are used in our adaptive transmission frameworks. It is important to mention that we propose two different transmission frameworks. The first transmission framework proposes a *bit*



Figure 7.1: Scheme of the OFDM system.

loading algorithm which minimizes the aggregate BER of an OFDM system while maintaining a target buffering delay of the packets in the transmission buffer. In this case, the transmit power, which is divided equally among the subcarriers is assumed to be fixed in each transmission slot. In the second transmission framework, we propose *bit and power* loading scheme which has the objective to minimize the average total transmit power while maintaining a target buffering delay of the packets.

7.1.1 Channel models

Let the total bandwidth of the system be W[Hz]. We assume that the total bandwidth is equally divided among L OFDM subcarriers. Therefore, the bandwidth per carrier is $W_{subcar} = W/L[\text{Hz}]$. Given the channel fading amplitude at *i*th carrier, α^i and a noise power spectral density of $N_0[W/\text{Hz}]$, let us define the normalized fading gain at receiver for the *i*th carrier $h^i = (\alpha^i)^2/N_0W_{subcar}$. The fading gain channel estimation for each subcarrier is performed at the receiver and these estimates are sent back to the transmitter assuming no delay and no error in the transmission on the feedback channel. The received fading gain estimate on carrier *i*, h^i corresponds to a discrete set $C^i = \{c_1^i, \ldots, c_Q^i\}$ of channel states where it is assumed that there are Q such discrete states. Let *i*th carrier be in the state c_j^i if h^i is in the interval $\phi_{j-1} \leq h^i \leq \phi_j$. Here it is assumed that fading gain thresholds satisfy the following inequality: $0 = \phi_0 \leq \ldots \leq \phi_j \leq \ldots \leq \phi_Q = \infty$. The channel is assumed to be block-fading as it is assumed that the received fading gain at the receiver stays in the given interval for the duration of the whole transmission slot. In each transmission slot, a given number of OFDM symbols are transmitted. The duration of a transmission slot, T_s , over which the fading gain remains roughly constant has to satisfy the following condition [74]

$$T_s << \frac{0.423}{f_d},$$
 (7.1)

where f_d is the maximum Doppler frequency which depends on the signal carrier frequency and the velocity of the mobile terminal. The PDF of the fading gain of *i*th carrier in state *l*, $p_l^i(h)$ is considered to be known. Further, it is assumed that channel states of *i*th carrier form an ergodic Markov chain which is independent across the carriers. The composite channel state is defined as $c = (c^1, c^2, \dots, c^L)$, where $c^i \in C^i$, defines the snapshot of carrier states in all parallel carriers. The number of composite channel states in state space $C = C^1 \times C^2 \times \cdots \times C^L$ is Q^L . If the fading conditions among the subcarriers are independent, they can be obtained by simply taking the product of individual carrier's state transition probability. For example, we assume that the channel of *i*th carrier can be represented by FSMC [71]. For a given fading rate (i.e., Doppler frequency) the transition probabilities between the channel states of the FSMC can be found, for example, in [71].

7.1.2 Incoming traffic model

The incoming traffic i.e., the packet arrival process at the transmission buffer is assumed to be constant \overline{F}_{sec} packets per second. If the duration of a slot is equal to $T_s[Sec]$, the packet arrival rate is $\overline{F} = \overline{F}_{sec}T_s$ packets per slot. It is important to mention that for a given total bandwidth, the symbol duration of an OFDM system depends on the number of sub-carriers. For example, if we assume a *L* carrier OFDM system and ideal Nyquist signalling, the duration of a symbol is $T_{sym} = LT_{sym}^{sc}$, where $T_{sym}^{sc} = 1/W$ is the duration of a symbol for the equivalent single carrier system. In our transmission frameworks we assume that N_s OFDM symbols are transmitted in each transmission slot irrespective of the number of carriers in the system. Therefore, the duration of a transmission slot, $T_s[Sec]$ is also different for different number of carriers in the system for a given total bandwidth. We also assume a packet length of N_p bits per packet. The number of packets that are transmitted in a slot depends on the total number of bits loaded in OFDM subcarriers in that given slot.

7.1.3 Adaptation model

The adaptive modulator in carrier *i* is assigned an action a^i from the set of possible actions \mathcal{A}^i . It is assumed that the same rate actions are available in all carriers. Without loss of generality we assume that the set of available actions $\mathcal{A}^i = \{a_0, a_1, a_2, a_3 \dots a_U\}, \forall i$ where action a_0 and a_u ($u = 1, 2, \dots, U$) correspond to transmitting no bit and u bits, respectively. The composite action space, \mathcal{A} is the combination of all possible transmission rates in each subcarrier and is defined as $a = (a^1, a^2 \dots, a^L)$, where $a^i \in \mathcal{A}^i$ and $a \in \mathcal{A}$ has total of $A = U^L$ actions. The role of the rate scheduler is to choose this composite action, $a \in \mathcal{A}$ based on the current composite channel state $c \in C$ and the current buffer state, $b \in \mathcal{B}$, where $\mathcal{B} = \{Z_1 Z_2 \cdots Z_{B+1}\}$ is the buffer state space. Taking an action based on the buffer as well as the composite channel state is specified by a decision rule which we call the policy. More specifically, the stationary deterministic policy φ , which does not depend on time index, specifies action $a_{(n)} = \varphi(b_{(n)}, c_{(n)})$.

If a feasible action¹⁹ $a_{(n)} = a \in \mathcal{A}$ is taken at time step n, the buffer dynamics can be expressed as

$$b_{(n+1)} = b_{(n)} + \bar{F} - \frac{N_s}{N_p} \sum_{i=1}^{L} \Psi^i(a_{(n)})$$
(7.2)

where $b_{(n+1)}$ and $b_{(n)}$ denote the buffer occupancy at time step n and (n + 1), respectively. $\Psi^i(a)$ returns the number of bits per symbol assigned to *i*th carrier if composite action $a_{(n)}$ is taken. This can be obtained by looking at the individual actions of *i*th carrier for that given composite action $a_{(n)}$. Without loss of generality, we use a normalized packet length i.e., the values of N_p and N_s in Eq. (7.2) are chosen such that the ratio N_s/N_p is equal to one. As a consequence we omit this ratio throughout the rest of the chapter. In practice, the packet length N_p can be thousands of bits per packet and the ratio can take any integer value.

First transmission framework

In the first transmission framework the total fixed transmit power, $P_{\tau}[W]$ is equally divided among the carriers. The fixed transmit power per carrier, therefore, is $P = P_{\tau}/L[W]$. The

¹⁹A feasible action does not lead to buffer overflow or negative buffer occupancy. Detailed description of feasible actions will be given in Section 7.2.1.

BER of a given subcarrier is generally a function of the CNR as well as the number of bits v assigned to the subcarrier and can be written as

$$BER(g,v) = f(g,v) \tag{7.3}$$

where g = hP is the CNR per carrier and $f(\cdot, \cdot)$ is the BER function determined by specific modulation and coding scheme. For example, when uncoded M-QAM is used in each subcarrier, the exact expression of this function can be found in [72] and a good approximation is expressed as (e.g., see [75])

$$f(g,v) = 0.2 \times \exp\left(\frac{-1.6g}{(2^v - 1)}\right).$$
 (7.4)

When binary-phase-shift (BPSK) modulation is used in each subcarrier, the BER is given as

$$BER_{BPSK} = \frac{1}{2} \operatorname{erfc}(\sqrt{g}). \tag{7.5}$$

Since we use discrete channel state which is mapped to a unique range of CNRs rather than a unique instantaneous CNR value, using the expression of average BER in a given channel state is appropriate. The average BER of *i*th carrier, $\overline{\text{BER}}$ over channel state c_j^i , $j = 1, 2, \dots, L$ can be written as

$$\overline{\text{BER}}(c_j^i, v) = \frac{\int_{\phi_j}^{\phi_{j+1}} f(\mathbf{P} * h^i, v) p_{h^i}(h^i) dh^i}{\int_{\phi_j}^{\phi_{j+1}} p_{h^i}(h^i) dh^i}.$$
(7.6)

Second transmission framework

In the second transmission framework, both the transmit power and bits are adaptively loaded in each subcarrier. At first we assume that there exist an ideal coding scheme that allows information transmission without any transmission error. Therefore, the transmit power, P^i corresponding to taking action a^i on *i*th carrier in channel state c_j^i is connected via the Shannon capacity formula and is given by

$$\mathbf{P}^{i}(c_{j}^{i},a^{i}) = \frac{1}{\bar{h}_{j}^{i}}(2^{a^{i}}-1)$$
(7.7)

where \bar{h}_{j}^{i} is the average received fading gain of *i*th carrier at channel state c_{j}^{i} . It is important to mention that in this case a^{i} represents the number of coded bits loaded on *i*th subcarrier. It is important to mention that in this transmission framework N_{s} is the number of coded symbols transmitted in each transmission slot. If the data packets can tolerate some transmission error instead of having an error free transmission, a practical modulation and coding scheme can be used. In this case the transmit power, P_{prac}^{i} corresponding to the rate action a^{i} can be approximated using the well known CNR gap approximation [76]

$$\mathbf{P}^{i}_{\text{prac}}(c^{i}_{j},a^{i}) = \frac{\Theta}{\overline{h}^{i}_{j}}(2^{a^{i}}-1),$$
(7.8)

where Θ is a coding and modulation dependent parameter. For example, with uncoded MQAM, Θ can be calculated as [76]

$$\Theta = \frac{1}{3} \left[Q^{-1} \left(\frac{\text{BER}_0}{4} \right) \right]^2 \tag{7.9}$$

where $Q^{-1}(\cdot)$ is the inverse standard Gaussian Q-function and BER₀ is the target BER.

7.2 Scheduling Algorithms

In this section we formulate our buffer and channel adaptive loading frameworks as the finite MDP. Then optimal bit-loading policies are obtained using equivalent LP methodology. Finally, a simple suboptimal algorithm which shows the performance very close to the optimal one is proposed.

7.2.1 **Problem formulation**

An unconstrained MDP is a model for deciding on how to act in a stochastic environment where outcomes are uncertain [73]. It is described through a set of decision epochs or time slots or stages $\mathcal{T} = 1, 2, \dots, n$, a set of states $\mathcal{S} = \{s^1, s^2, \dots, s^N\}$, a set of actions $\mathcal{A} = \{a^1, a^2, \dots, a^U\}$, a set of state and action dependent immediate costs $x: \mathcal{K} \mapsto \mathbb{R}$, and a set of state and action dependent transition probabilities. Set $\mathcal{K} = \{(s, a) : s \in \mathcal{S}, a \in \mathcal{A}_s\}$ is the set of state-action pairs. While the system is in state $s_{(n)} = s^j \in \mathcal{S}$ at decision epoch or time-slot n, the decision maker selects an action $a_{(n)}$ from the set of actions \mathcal{A}_{s^j} available at state $s^j \in \mathcal{S}$, where $\bigcup_{s^j \in \mathcal{S}} \mathcal{A}_{s^j} = \mathcal{A}$. In the next time-slot, the system moves to a state $s^k \in \mathcal{S}$ according to the transition probability $p(s_{(n+1)} = s^k | s_{(n)} = s^j, a_{(n)})$ and the decision maker incurs a one step immediate cost $x(s_{(n)}, a_{(n)})$. A policy as a mapping from state to actions specifies the decision rule to be used at all time slots. With a stationary deterministic policy φ , the expected long-term average cost per stage is

$$J_{\varphi} = \lim_{M \to \infty} \frac{1}{M} \sum_{n=1}^{M} \mathsf{E}_{s}[x(s_{(n)}, \varphi(s_{(n)}))],$$
(7.10)

where expectation, $E_s[\cdot]$ is over random evolution of the states of the MDP. An optimal policy denoted by φ^* is the one that incurs the minimum average cost per stage i.e.,

$$\varphi^* = \arg\min_{\varphi \in \Pi} J_{\varphi} \tag{7.11}$$

where Π is a set of all stationary policies.

The composite state space S of adaptive loading framework is composed of the buffer state space B and composite channel state C. The number of states in $S = C \times B$ is $N = (B + 1) \times Q^L$. Since a finite size buffer is used at the transmitter, taking certain actions at specific buffer states leads to buffer overflow. This overflow can be avoided by taking an action $a_{(n)}$ at time slot n such that

$$b_{(n)} + \bar{F} - \sum_{i=1}^{L} \Psi^{i}(a_{(n)}) \le B.$$
 (7.12)

On the other hand, the transmitter cannot transmit more packets than available at the buffer. So, the action a has to follow the inequality

$$\sum_{i=1}^{L} \Psi^{i}(a_{(n)}) \le b_{(n)}.$$
(7.13)

Using Eqs. (7.12) and (7.13) a feasible action space in state s, A_s , where, $\bigcup_{s \in S} A_s = A$ can be found easily.

Transition probabilities between states are based on transition probabilities of channel Markov chains and buffer dynamics Eq. (7.2) and can be expressed as follows:

$$p(s_{(n+1)} = s^{j}|s_{(n)} = s^{i}, a_{(n)}) = \operatorname{Prob.}(\Lambda(s^{j})|\Lambda(s^{i}))\delta(\nabla(s^{j}) - (\nabla(s^{i}) - \sum_{k=1}^{L} \Psi^{k}(a_{(n)}) + \bar{F}))$$
(7.14)

for certain $a_{(n)} = a \in \mathcal{A}_{s^i}$, and s^i , $s^j \in S$. Functions $\Lambda(s)$ and $\nabla(s)$ return the composite channel state c and the buffer state b of composite state s, respectively. $\delta(x)$ returns 1 if x = 0 and 0 otherwise.

First transmission framework

In our proposed bit-loading framework which minimizes the aggregate BER with a fixed transmit power, we have two types of cost; delay cost and aggregate BER cost. The aggregate BER of the multicarrier system in state, $s_{(n)} = s \in S$ for taking an action $a_{(n)} = a \in A_{s_n}$ can be written as

$$BER_{agg}(s_{(n)}, a_{(n)}) = \frac{\sum_{i=1}^{L} \overline{BER}(\ell^{i}(s_{(n)}), \Psi^{i}(a_{(n)})) \Psi^{i}(a_{(n)})}{\sum_{i=1}^{L} \Psi^{i}(a_{(n)})},$$
(7.15)

where ℓ^i returns *i*th carrier channel state of composite state *s*. The expected long-term average BER cost under policy φ is

$$\overline{\text{BER}}_{\text{agg}}(\varphi) = \lim_{M \to \infty} \frac{1}{M} \sum_{n=1}^{M} \mathsf{E}_{s}[\text{BER}_{\text{agg}}(s_{(n)}, \varphi(s_{(n)}))].$$
(7.16)

In order to avoid notational confusion, it should be mentioned that, in general, BER_{agg} and $\overline{\text{BER}}_{agg}$ denote the immediate and the long term average aggregate BER, respectively. The delay cost is the average number of time slots that a packet has to wait at the transmission buffer before transmission. The immediate delay cost due to buffer occupancy under the policy φ is

$$d(s_{(n)}, \varphi(s_{(n)})) = \frac{\nabla(s_{(n)})}{\bar{F}}.$$
(7.17)

According to Little's formula, the time average delay in the buffer, D can be expressed as

$$\bar{\mathbf{D}}(\varphi) = \lim_{M \to \infty} \frac{1}{M} \sum_{n=1}^{M} \mathsf{E}_{s}[\frac{b_{(n)}}{\bar{F}}].$$
(7.18)

Since both the costs; BER and delay costs are conflicting with each other, we can formulate the problem as an unconstrained Markov decision process (UMDP) with a trade-off factor or a constrained Markov decision process (CMDP). When the problem is posed as an UMDP, the immediate cost function under a stationary deterministic policy φ is

$$x(s_{(n)},\varphi(s_{(n)})) = \frac{\nabla(s_{(n)})}{\bar{F}} + \beta \text{BER}_{agg}(s_{(n)},\varphi(s_{(n)}))$$
(7.19)

where β (> 0) is a trade-off factor and by changing its value we can obtain different tradeoffs between average delay and aggregate BER. For a given β , the stationary deterministic optimal policy φ^* can be obtained using dynamic programming such as RVI or PI [73]. Minimum aggregate BER and average delay under the policy φ^* are obtained from Eqs. (7.16) and (7.18).

In CMDP, the objective cost is minimized while keeping the additional costs (called constrained costs) below some given bounds. In this proposed framework, our objective is to minimize the aggregate BER while keeping delay cost below a target threshold, the problem thus can be formulated as follows:

$$\arg\min_{\varphi \in \Pi_R} \overline{\text{BER}}_{\text{agg}}(\varphi) \tag{7.20}$$

subject to:
$$\overline{D}(\varphi) \le D_t$$
 (7.21)

where Π_R is the set of all randomized stationary policies that are uniquely characterized with probabilities $\theta_{(s)}^{\varphi}(a)$ of applying action $a \in \mathcal{A}_s$ in state $s \in \mathcal{S}$. D_t is an allowable maximum long-term average delay. The formulated CMDP can be solved using equivalent LP methodology [73]. A brief description on equivalent LP methodology to find optimal bit-loading policies is given in the next Section.

Second transmission framework

Similarly, in a bit and power loading framework which minimizes transmit power while achieving a target delay constraint, there are two types of costs. One is buffering delay cost as described in Eq. (7.17) and the other is power cost which can be expressed as

$$\mathbf{P}_{\mathsf{T}}(s_{(n)}, a_{(n)}) = \sum_{i=1}^{L} \mathbf{P}^{i}(\ell^{i}(s_{(n)}), \Psi^{i}(a_{(n)}))$$
(7.22)

where $P^i(c^i, a^i)$, in general, is the required transmit power in *i*th carrier for taking an action a^i at state c^i which can be obtained from Eq. (7.7) or (7.8). The expected long term average total power cost under policy φ is

$$\overline{\mathsf{P}}_{\mathsf{T}}(\varphi) = \lim_{M \to \infty} \frac{1}{M} \sum_{n=1}^{M} \mathsf{E}_{s}[\mathsf{P}_{\mathsf{T}}(s_{(n)}, \varphi(s_{(n)}))].$$
(7.23)

Again, in case of minimizing the total transmit power while meeting a given delay constraint, we have two options to pose the optimization problem at hand. One is formulating the problem as an UMDP process and the other is formulating the problem as a CMDP. When the problem is posed as an UMDP, the immediate cost function under a stationary deterministic policy φ is

$$x(s_{(n)},\varphi(s_{(n)})) = \frac{\nabla(s_{(n)})}{\bar{F}} + \beta \mathsf{P}_{\mathsf{T}}(s_{(n)},\varphi(s_{(n)})).$$
(7.24)

With a CMDP the power optimization problem can be formulated as follows:

$$\arg\min_{\varphi\in\Pi_R}\overline{\mathsf{P}}_{\mathsf{T}}(\varphi) \tag{7.25}$$

subject to:
$$\overline{D}(\varphi) \le D_t$$
. (7.26)

7.2.2 Optimal scheduling

Although dynamic programming offers an elegant, unified treatment of a wide range of stochastic control problems, the curse of dimensionality gives rise to prohibitive computational requirements of large-scale problems, such as this OFDM bit loading problem. As we mentioned in previous section, there are number of algorithms, for example, RVI, PI, and LP that can be used to find the optimal solution for stochastic optimization problems formulated as a MDP. LP is studied in solving MDP models because of its elegant theory and the ease in which it allows inclusion of constraints [73]. To describe the dual LP²⁰ let us consider that z(s, a) represents the "steady-state" probability that the process is in state s and action a is applied. We seek to find the control policy which is represented in terms of probability distribution z over $S \times A$. The optimal policy z^* can be obtained as

minimize
$$\sum_{s \in S, a \in A_s} \text{BER}_{agg}(s, a) z(s, a)$$
 (7.27)

subject to :

$$\sum_{s \in \mathcal{S}, a \in \mathcal{A}_s} \mathsf{d}(s, a) z(s, a) \le D_t \tag{7.28}$$

$$\sum_{a \in \mathcal{A}_j} z(j, a) = \sum_{s \in \mathcal{S}, a \in \mathcal{A}_s} z(s, a) p(j|s, a), \ \forall j \in \mathcal{S}$$
(7.29)

$$\sum_{s \in \mathcal{S}, a \in \mathcal{A}_s} z(s, a) = 1 \tag{7.30}$$

$$z(s,a) \ge 0, \forall s \in \mathcal{S}, \forall a \in \mathcal{A}_s.$$
(7.31)

²⁰In order to describe the LP methodology, we consider the problem of minimizing aggregate BER as an example. However, the transmit power minimization problem can be considered using the same methodology.

The inequality constraint in Eq. (7.28) is to keep the long-term average delay below the given threshold whereas the equality constraint in Eq. (7.29) satisfies the well known Chapman-Kolmogorov equation. The constraint in Eq. (7.30) ensures that the sum of the probabilities z(s, a) is equal to one while Eq. (7.31) is ensuring that the individual probability is greater than zero.

If there exist an optimal solution z^* to the LP, there exists an randomized stationary policy φ^* that is optimal for the CMDP. In general, the optimal policy φ^* for a CMDP is randomized. Now

$$\theta_{(s)}^{\varphi^*}(a) = \frac{z^*(s,a)}{\sum_{a' \in \mathcal{A}_s} z^*(s,a')} \quad \text{if } \sum_{a' \in \mathcal{A}_s} z^*(s,a') > 0.$$
(7.32)

If $\sum_{a' \in A_s} z^*(s, a') = 0$ for some $s \in S$, an action that drives the system to $S_{z^*} = \{s \in S : \sum_{a' \in A_s} z^*(s, a') > 0\}$ is chosen in each state [73].

The delay-BER trade-offs for different number of carriers are shown in Fig. 7.2. In order to make a fair comparison for different number of carriers, we used the same total transmit normalized (w.r.t. N_0W) power, $P_T = 1$. Due to the high complexity of finding optimal policies we had to restrict our simulations up to four carriers. We used three state Rayleigh channel model for each carrier. The states are partitioned such that probabilities of being in each state are equal. For example, this type of channel model can be found in [65], [66]. The channel variations across the subcarriers are assumed to be identically and independently distributed (i.i.d) with an average received power gain, $h_0 = 15$ [dB]. The available rate set per carrier, $\mathcal{A}^i = [0 \ 2 \ 4]$. A packet arrival rate of 2 packets per time slot for equivalent single carrier case is used and thus the packet arrival rate for k carrier system is $2 \times k$ packets per time slot. As mentioned in Section 7.1.2 the slot duration of an OFDM system with k subcarriers is $T_s^k = k \times T_s^{sc}$, where T_s^{sc} is the slot duration of an equivalent single carrier system. Therefore, in Fig. 7.2, the average delay for different number of carriers is normalized by the duration T_s^{sc} . In our proposed framework, the minimum achievable delay is one time slot. The normalized minimum achievable delay is, therefore, k times that of an equivalent single carrier system and consequently the minimum achievable delay starts from k time slots rather than starting from one time slot. This figure also shows that as the maximum tolerable delay increases, the aggregate BER decreases. This is expected because of temporal diversity (TD) effects in time domain. The packet can wait at the transmission buffer until the channel condition is good enough to be transmitted. If



Figure 7.2: Optimal delay and aggregate BER trade-offs with different number of carriers (Doppler frequency, $f_d = 300$ Hz).

there is no delay constraint, the minimum achievable aggregate BER, which corresponds to water filling case, can be achieved. It is interesting to see that as the number of carriers increases in the system, the delay limited aggregate BER, in high delay tolerant region, decreases for the same traffic arrival rate (packets per second). This is owing to the spectral diversity (SD) effects in frequency domain. As there are more carriers in the system, more SD is achieved. It is also interesting to observe that in low delay-constrained region, although the system has higher number of carriers, the aggregate BER is higher (see crossover points between the curves in Fig. 7.2). This is due to the fact that in that region, the TD takes over the SD. In other words, the TD is more dominant than SD.

The delay-power tradeoffs for different number of carriers are shown in Fig. 7.3. In this case we assume ideal coding scheme with rate set per carrier, $\mathcal{A}^i = [0 \ 2 \ 4]$. Therefore, the corresponding transmit normalized power levels were obtained using Eq. (7.7). Fig. 7.3 shows that as the maximum tolerable delay increases, the total transmit power decreases as expected due to TD as explained above. Another important remark from the figure is that as the number of carriers increases in the system, the SD gain becomes marginal. In other



Figure 7.3: Optimal delay and power tradeoffs with different number of carriers (Doppler frequency, $f_d = 100$ Hz).

words, as we add more carriers in the system, the power saving that results from the extra additional carrier decreases. For example, when the system has two carriers instead of one, it saves a lot more power in order to meet the same target delay. However, adding the third carrier does not save that much power and the same scenario extends to the fourth carrier.

7.2.3 Suboptimal algorithm

The complexity of finding the optimal bit-loading policy is enormous for large number of parallel carriers as it is often the case in OFDM systems. This phenomenon is commonly referred to as the curse of dimensionality. Even if we could solve the problem using some advanced software which can handle LP with millions of variables, such a policy might not be realizable practically because of memory requirements of the lookup table to store the optimal policies. Therefore, we resort to the suboptimal solution. The principle is based on the divide and conquer approach. Optimal rate scheduling algorithm is produced for the equivalent single carrier system and then, using this solution, transmission rate for each

carrier is found considering a carrier's state and buffer state individually. The sum rate, i.e., the summation of these individual carrier rates is distributed among the carriers using a greedy approach in order to find the final suboptimal rate allocation policy. The *suboptimal rate allocation algorithm* is described below step-by-step

- Step 1: Find the optimal transmission rate a^i for each carrier assuming that the carriers are independent
- Step 2: Find the sum rate $R_{sum} = \sum_{i=1}^{L} a^i$
- Step 3: Divide the sum rate R_{sum} among the carriers using the greedy approach, according to the composite channel state c.

In Step 1, the optimal transmission rate for each carrier is found from the optimal policy set of an equivalent single carrier system with a finite transmission buffer of size B and the system bandwidth of W[Hz]. The stationary optimal deterministic policy of the single carrier transmission system can be found formulating the problem as an UMDP (see, for example, [66]). It should be mentioned that as the total number of states is much less than in the multicarrier system, RVI can be used to find the optimum deterministic policy. Once the optimal policy for single carrier system for a fixed tradeoff factor β is available, the total sum rate is found adding the individual optimal rates as follows:

- Initialize the set \mathcal{L} to the set of all carriers.
- Pick up the best carrier i from a set of all unassigned carriers L. Without loss of generality, consider that this carrier is in state c_jⁱ, and the buffer being in state b_(n). The composite state for this carrier is s_{sc} = (b_(n), c_jⁱ). The subscript "sc" is to mean the composite state of the equivalent single carrier case.
- Find the optimal action a*(s_{sc}) corresponding to the state s_{sc} from the Table of equivalent single carrier case optimal policy. Update the buffer occupancy b_(n) = b_(n) a*(s_{sc}) and remove the carrier i from the set of unassigned carriers L.
- Repeat the above procedure until all the carriers are assigned.
- Add the transmission rate in all subcarriers to find the sum rate R_{sum} in Step 2 of the suboptimal rate allocation algorithm.



Figure 7.4: Optimal and sub-optimal delay and aggregate BER trade-offs for different fading rates in two carrier case.

In step 3, the greedy allocation, which has been used in literature for optimal resource allocation (see, for example, [69], [68]) improves the rate allocation among the carriers. The greedy rate allocation procedure will be described in the next section.

The delay-BER tradeoffs with the proposed suboptimal bit-loading algorithm with different number of carriers are plotted in Figs. 7.4, 7.5 and 7.6, respectively by changing the tradeoff factor β of a single carrier policy. We have also plotted the optimal delay-BER tradeoffs in Figs. 7.4 and 7.5. In these examples, we have considered three channel states per carrier and a rate set per carrier is $a^i = [0\ 1\ 2\ 3\ 4]$. In suboptimal policy performance simulations, due to the MATLAB software-based simulation time requirements we had to restrict ourselves up to a four carrier OFDM system. In a real system, there are two options to implement the suboptimal loading scheme. One way is to store the single carrier system optimal policies, which are obtained off-line, in a lookup table. These pre-stored single carrier optimal policies are then used to find the suboptimal policy for each carrier using the greedy allocation, in online fashion, as described above. Since one needs to store only single carrier system's policy, it requires less memory space but requires more online com-



Figure 7.5: Optimal and sub-optimal delay and aggregate BER trade-offs for different fading rates in three carrier case.

putation. The other option is to find the suboptimal policies for multicarrier system using the single carrier system optimal policies and the greedy allocation off-line. Then these suboptimal policies are stored in a lookup table. This strategy requires less on-line computation but requires more memory space. Figs. 7.4, 7.5 and 7.6, show that as the fading rate increases (i.e., the channel fading rate becomes faster), the aggregate BER decreases in order to maintain the same delay constraint. This is due to the influence of the fading rate on TD. As the channel becomes faster, the probability that the packets will experience a better channel condition within a given number of transmission slots increases. In this regard, it is also worthy to mention that the minimum achievable aggregate BER in order to maintain a hard delay constraint (a delay of one time slot), is not affected by the fading rate. In this hard delay-constrained case since the packets are to be transmitted within one time slot, TD is not achieved at all. The figures also show that the sub-optimal algorithm offers almost the same performance as the optimal one and as the number of carriers in the system increases, the performance of the suboptimal algorithm slightly degrades compared with the optimal algorithm.



Figure 7.6: Sub-optimal delay and aggregate BER trade-offs for different fading rates in four carrier case.



Figure 7.7: Optimal and sub-optimal delay and power trade-offs for different fading rates in two carrier case.

Similarly we have plotted the suboptimal delay-power tradeoffs in Figs. 7.7, 7.8 and 7.9. As expected with the increase of the fading rate, the total average transmit power decreases in order to maintain the same delay constraint. The power-delay trade-off curves for different fading rates start from the same point which is expected in the hard delay-constrained case. These figures also show that the proposed suboptimal policy attain delay-power tradeoffs which are very close to the optimal ones. One further effective method to reduce the complexity of OFDM loading problem is block wise loading (see, e.g. [68]). In this block loading method, it is assumed that the neighboring carriers undergo roughly the same fading gain. So, the neighboring carriers are grouped in a block and each block is loaded with the same rate (or both power and rate) which reduces the complexity of the loading algorithm compared with loading each carrier individually. For example, an OFDM system with 48 carries will have 12 such blocks if 4 neighboring carriers are grouped together to form a block. If we use the block wise loading methodology, the number of composite states as well as total number of actions will be reduced and eventually it will lead to a simpler implementation.



Figure 7.8: Optimal and sub-optimal delay and power trade-offs for different fading rates in three carrier case.



Figure 7.9: Sub-optimal delay and power trade-offs for different fading rates in four carrier case.

7.3 Effect of Correlation among Subcarriers' Fading Gains

In this section, we study and explore the effect of correlation between the subcarriers' fading gains of a MCM system on its bit and power loading for delay-constrained and no delay-constrained communication. We consider both *continuous* and *discrete* rate and channel gain variation cases. For continuous channel gain and rate variation, Lagrange formulation is used in order to find the optimal loading policy. For *discrete rate adaptation* and *discrete channel gain variation*, the Lagrangian formulation is not applicable. Therefore the optimal loading policies for both delay-constrained and no delay-constrained cases are obtained via a greedy allocation.

7.3.1 Channel and rate allocation model

The composite channel gains of all subcarriers which are assumed to be known at the transmitter, are denoted by a vector, $\mathbf{h} = (h^1, h^2, \dots, h^L)$. For continuous channel model $h^i \in \Re^+, i = 1, \dots, Q$ and the set of possible composite channel gains is $\mathbf{h} \in \mathcal{H} = (\Re^+)^Q$ with joint PDF of the composite channel state denoted by $p_{\mathbf{h}}(\mathbf{h})$. For discrete channel model $h^i \in \{h_1, \dots, h_Q\}, i = 1, \dots, L$ and the composite channel gain \mathbf{h} of all the carriers belongs to the set $\mathcal{H} = \{\mathbf{h}_1, \dots, \mathbf{h}_{H^L}\}$ where $\mathbf{h}_i, i = 1, \dots, H^L$, are all combinations of discrete channel states of all parallel carriers with probability of $\mathbf{h} \in \mathcal{H}$ denoted by $\mathbb{P}(\mathbf{h})$.

In each transmission slot, the transmitter adaptively assigns the transmission rate of kth carrier, R_k and adjusts the transmit power level, P_k by joint consideration of the different subcarriers' received power gains i.e., h. Therefore let us consider the transmission rate of kth carrier is function of h and the transmission rates in the parallel subcarriers denoted by the vector, $\mathbf{R} = (R_1(\mathbf{h}), R_2(\mathbf{h}), \dots, R_L(\mathbf{h}))$. In continuous rate adaptation it is assumed that any positive rate is achievable i.e. $R^i(\mathbf{h}) \in \Re^+$, whereas in the discrete rate case it is assumed that $R^i(\mathbf{h}) \in \{0, 1, 2, ...\}$ for all i = 1, ..., L. In order to adjust the transmit power we use the well known SNR gap approximation in (7.8).

7.3.2 Rate allocation with no delay constraints

Water-filling power, P_{wF} defined as the minimum required average transmit power of a MCM system with no delay constraint can be expressed as

$$\mathbf{P}_{\mathsf{WF}} = \min_{R_i(\mathbf{h})} \mathsf{E}\left[\sum_{i=1}^{L} \mathsf{P}^i(R^i(\mathbf{h}), h^i))\right], \tag{7.33}$$

s.t.
$$\mathsf{E}\left[\sum_{i=1}^{L} R^{i}(\mathbf{h})\right] \ge \bar{F}$$
 (7.34)

where the operator $E[\cdot]$ is an expectation operator taken with respect to the joint PDF $p_h(h)$. In Eq. (7.34), the inequality constraint is to ensure that the average sum transmission rate in the parallel carriers is greater or equal to the data rate \overline{F} . The problem of maximizing average transmission rate (i.e., water-filling capacity) with an average power constraint, which is the dual problem of our optimization problem in Eq. (7.33), has been solved earlier [75].

Continuous rate and channel variation: Using Eqs. (7.33) and (7.34), we form the following Lagrangian

$$J(\mathbf{R}(\mathbf{h}), \hat{\lambda}) = \mathsf{E}\left[\sum_{i=1}^{L} \mathsf{P}^{i}(R^{i}(\mathbf{h}), h^{i})\right] + \hat{\lambda}\mathsf{E}\left[\sum_{i=1}^{L} R^{i}(\mathbf{h})\right],\tag{7.35}$$

where $\hat{\lambda}$ is the Lagrange multiplier. Finding the gradient of (7.35) with respect to $\mathbf{R}(\mathbf{h})$ and equating it to zero, we get L independent equations in $R_k(\mathbf{h})$ that solve to

$$R^{i}(\mathbf{h}) = R^{i}(h^{i}) = W \log_{2} \left(-\frac{\hat{\lambda}Wh^{i}}{\Theta \ln 2}\right).$$
(7.36)

The expression for the rate allocation in Eq. (7.36) is also optimal for the single carrier (L = 1) case for some $\hat{\lambda}$. Lagrangian constraint $\hat{\lambda}$ can now be obtained from the constraint on the average sum rate in Eq. (7.34)

$$\sum_{i=1}^{L} \int_{h^{1}=0}^{\infty} \cdots \int_{h^{L}=0}^{\infty} R^{i}(\mathbf{h}) p_{\mathbf{h}}(\mathbf{h}) dh^{1} \cdots dh^{L} = \bar{F}$$
(7.37)

which employing $R^{i}(\mathbf{h}) = R^{i}(h^{i})$ simplifies to

$$\sum_{i=1}^{L} \int_{h^{i}=0}^{\infty} R^{i}(h^{i})p(h^{i})dh^{i} = \bar{F}$$
(7.38)

149

where $p(h^i)$ is the marginal distribution of the received power gain h^i , i = 1, ..., L. Eqs. (7.36) and (7.38) show that the loading policy in no delay-constrained case with continuous rate adaptation decouples into L single carrier rate allocation problems that is *dependent* only on the marginal distributions of channel gains in each channel. In other words, the water-filling power P_{WF} does not depend on the correlation among the subcarriers' fading gains. Intuitively, the result is also expected. In no delay-constrained case a data bit can wait theoretically infinite amount of time. Therefore, full diversity is achieved in both domains; SD and TD and it does not matter whether the channel fading gains are correlated or not.

Discrete rate and channel variation: Let the discrete rates and discrete received power gains be defined as in Section 7.3.1. In this case the optimal rate allocation of problem defined in Eqs. (7.33) and (7.34) can be solved using the greedy allocation. It is important to mention that the Lagrangian formulation is not applicable to discrete rate variation case. The greedy allocation is performed both in time and spectral domain and is given as follows:

• Step 1: Form a discrete rate allocation vector $\hat{\mathbf{R}}$ with $H = Q \times H^L$ elements

$$\hat{\mathbf{R}} = \{R^1(\mathbf{h}_1), \dots, R^1(\mathbf{h}_{H^L}), \dots, R^L(\mathbf{h}_1), \dots, R^L(\mathbf{h}_{H^L})\}$$

and set $\hat{\mathbf{R}} = \{0, 0..., 0\}$. Let

$$\hat{\mathbf{P}}(\hat{\mathbf{R}}) = \sum_{i=1}^{L} \sum_{l=1}^{H^{L}} \mathbb{P}(\mathbf{h}) \mathbf{P}(R^{i}(\mathbf{h}_{l}), \mathbf{h}_{l}(i))$$
(7.39)

be the average power needed for the discrete rate allocation vector $\hat{\mathbf{R}}$, where $\mathbf{h}_l(i)$ is the normalized power gain of carrier *i*. of composite state \mathbf{h}_l .

• Step 2: Let $\hat{\mathbf{R}}_{l} = \{R(1), \dots, R(l) + 1, \dots, R(H)\}$, for $l = 1, \dots, H$. Find \hat{l} such that

$$\hat{l} = \arg\min\left(\hat{P}(\hat{\mathbf{R}}_l) - \hat{P}(\hat{\mathbf{R}})\right)$$
(7.40)

and set $\mathbf{R} = \hat{\mathbf{R}}_{\hat{i}}$.

• Step 3: If (7.34) is satisfied stop; else go to step 2.

As a numerical example we consider a two carrier MCM system whose subcarriers' gains are faded according to bivariate Rayleigh distribution. The bivariate Rayleigh distribution can be expressed as (see, for example, [77])

$$= \frac{p_{(\alpha_1,\alpha_2)}(\alpha_1,\Omega_1;\alpha_2,\Omega_2|\rho)}{\Omega_1\Omega_2(1-\rho)} \exp\left[-\frac{1}{1-\rho}\left(\frac{\alpha_1^2}{\Omega_1}+\frac{\alpha_2^2}{\Omega_2}\right)\right] \cdot I_0\left[\frac{2\sqrt{\rho}\alpha_1\alpha_2}{(1-\rho)\sqrt{\Omega_1\Omega_2}}\right]$$
(7.41)

where $\alpha_1, \alpha_2 \ge 0$, $\Omega_k = \mathsf{E}[\alpha_k^2]$, k = 1, 2, and $\rho = \operatorname{cov}(\alpha_1^2, \alpha_2^2)/(\operatorname{var}(\alpha_1^2)\operatorname{var}(\alpha_2^2))$ is the correlation coefficient and $0 \le \rho < 1$ and $I_0(\cdot)$ is the modified bessel function of first kind and zero order. In our numerical example we used the same channel gain variance of $\operatorname{var}(|\alpha_k|^2) = 1$ in each carrier, $W_{subcar} = 1[\mathsf{KHz}]$ and $\Theta N_0 = 10^{-6}[\mathsf{W/Hz}]$. To simulate the discrete channel model, two dimensional space of bivariate Rayleigh distribution is quantized based on a uniform grid with 10 levels with step size 0.2. For discrete rate adaptation and discrete channel variations, the water-filling power P_{w_F} for two carriers system is shown in Fig. 7.10. In this plot, we have used two different values of ρ ; $\rho = 0.1$ and $\rho = 0.8$ and it shows that for a given average rate constraint there is an insignificant difference between the values of P_{w_F} for different values of ρ . This very small difference comes from the fact of discretization can be decoupled into L single carrier loading problems or not. However, numerically it is observed that for larger number of discrete channel states, the loading policy does not change with different values of ρ .

7.3.3 Rate allocation with hard delay constraints

Delay-limited power, P_{DL} , defined as the minimum required average total transmit power of a MCM system with hard delay constraint can be written as

$$\mathbf{P}_{\mathsf{DL}} = \min_{R^{i}(\mathbf{h}), i=1, \dots, L} \mathsf{E}\left[\sum_{i=1}^{L} \mathsf{P}(R^{i}(\mathbf{h}), h^{i})\right], \qquad (7.42)$$

s.t.
$$\sum_{i=1}^{L} R^{i}(\mathbf{h}) \ge \bar{F},$$
 (7.43)

In Eq. (7.43) the inequality constraint means that the instantaneous sum-rate in all parallel carriers should be greater than or equal to the data rate in order to meet the hard delay constraint. Maximization of sum-capacity of MCM system with an instantaneous power



Figure 7.10: Power versus rate constraint for delay limited and non-delay limited rate allocation.

constraint, which is a dual problem of the optimization in Eqs. (7.42) and (7.43), has been solved in [67]. For discrete rate adaptation, the optimal rate allocation for a given composite state $\mathbf{h}_i \in \mathcal{H}$ can be obtained using the greedy allocation of Section 7.3.2 by using $\mathbb{P}(\mathbf{h}) = 1$ if $\mathbf{h} = \mathbf{h}_i$ and $\mathbb{P}(\mathbf{h}) = 0$ if $\mathbf{h} \neq \mathbf{h}_i$ (see also [68]).

For the numerical settings of Section 7.3.2, the hard delay limited power, P_{DL} is plotted in Fig. 7.10 for different values of ρ . This figure shows that as the correlation between subcarriers' fading gains increases, the value of P_{DL} increases. This phenomenon can be explained as follows. In hard delay limited case since the data transmission rate is to be guaranteed at a fixed rate \overline{F} in each time slot, a data bit cannot wait in order to achieve TD in time domain. However, due to having more than one carrier in the system, SD among the carriers in frequency domain can be achieved. If the fading gains among the carriers are correlated, less SD is achieved and consequently required transmit power increases in order to maintain the same instantaneous transmission rate. It is worthy to mention that full SD is achieved if the subcarriers' fading gains are independently distributed and the required power in this case is the lower bound of P_{DL} .

7.4 Conclusions

In this chapter, we have studied an optimal and suboptimal loading algorithms of an OFDM system for delay-constrained communication over correlated fading channels. Specifically, we have proposed two different transmission frameworks for delay-constrained OFDM communication. One is the bit-loading framework which minimizes aggregate BER of a constant power OFDM system while meeting a delay constraint of the packets at the transmission buffer. The other is the bit and power loading algorithm which has the objective to minimize the total average transmit power while maintaining a target delay constraint. Optimal loading policies in both cases are obtained using LP methodology formulating them as the MDP. Since finding the optimal policies and implementing them in reality are highly complex, if not impossible, we have also explored a suboptimal loading algorithm based on an equivalent single carrier optimal loading scheme and a greedy approach. Since the suboptimal loading algorithm requires only finding and storing the optimal policy of the equivalent single carrier case with fewer number of states, it is simple to realize it practically. The selected numerical results have shown that the suboptimal loading scheme in both cases attains a very close tradeoff performance as the optimal one.

Finally, we have explored the effect of correlation between the subcarriers' fading gains of a MCM system on its bit and power loading in two extreme cases: delay-constrained and no delay-constrained communications. For no delay-constrained case the correlation between subcarriers fading gains does not have any influence on the loading policy. Therefore, the average total transmit power is not influenced by the correlation between between the subcarriers fading. For discrete rate adaptation as well as discrete channel gain variation, the optimal loading policies for both cases have been obtained via a greedy allocation. Assuming discrete rate adaptation and discrete channel gain variation, numerical examples have been demonstrated for different degrees of correlation. The results have showed that the required transmit power in hard delay-constrained case increases as the correlation increases. However, the degrees of correlation does not have any significant effect on the required transmit power in no delay-constrained case. Although we have considered only correlation among the carriers' fading gains, it is worthy to mention that time correlation does not influence our conclusions.

.

Bibliography

- [63] S. Shakkottai, T. S. Rappaport, and P. C. Karlsson, "Cross-layer Design for Wireless Networks," *Submitted for possible Journal publication*, June 2003. Also available in http://www.ece.utexas.edu/ shakkott/Pubs/cross-layer.pdf.
- [64] R. Berry and R. Gallager, "Communication over fading channels with delay constraints," *IEEE Trans. Inform. Theory*, vol. 48, pp.1135-1149, May 2002.
- [65] A. T. Hoang and M. Motani, "Buffer and channel adaptive modulation for transmission over fading channels," in *Proc. IEEE Int. Conf. Commun. (ICC'03)*, Anchorage, AK, May 2003, pp. 2748-2752.
- [66] D. Djonin, A. K. Karmokar, and V. K. Bhargava, "Optimal and sub-optimal packet scheduling over correlated fading channels," in *Proc. IEEE Int. Conf. Commun.* (*ICC'04*), Paris, France, Jun. 2004, pp. 906-910.
- [67] L. Goldfeld and V. Lyandres, 'Capacity of the multicarrier channel with frequencyselective Nakagami fading," *IEICE Trans. Commun.*, vol. E83-B, pp. 697-702, Mar. 2000.
- [68] C. Y. Wong, R. S. Cheng, K. B. Letaief, and R. D. Murch, "Multiuser OFDM with adaptive subcarrier, bit, power, and power allocation," *IEEE J. Select. Areas Commun.*, vol. 17, pp. 1747-1758, Oct. 1999.
- [69] D. N. C. Tse and S. V. Hanly, "Multiple access fading channels-Part I: Polymatriod structure, optimal resource allocation and throughput capacities", *IEEE Trans. Inform. Theory*, vol. 44 no. 7, pp. 2796-2815, November 1998.
- [70] L. Yang and M.-S. Alouini, "Cross layer analysis of buffered adaptive multicarrier transmission," in *Proc. IEEE Veh. Technol. Conf. (VTC'04 Fall)*, Los Angles, CA, September 2004.

- [71] H. S. Wang and N. Moayeri, "Finite-state Markov channel-A useful model for radio communication channels," *IEEE Trans. Veh. Technol.*, vol.44, pp. 163-171, Feb. 1995.
- [72] P. K. Vitthaladevuni and M.-S. Alouini, "A recursive algorithm for the exact BER computation of generalized hierarchical QAM constellations," *IEEE Trans. Inform. Theory*, vol. 49, no. 1, pp. 297-307, January 2003.
- [73] M. L. Putterman, "Markov decision processes : discrete stochastic dynamic programming," New York : John Wiley & Sons, 1994.
- [74] T. S. Rappaport, "Wireless Communications Principles and Practice," Prentice Hall, NJ, 2002.
- [75] S. T. Chung and A. J. Goldsmith, "Adaptive multicarrier modulation for wireless systems," in *Proc. The thirty-fourth Asilomar Conf. Signals, Systems and Comp.*, vol. 2, 29 Oct.-1 Nov. 2000, pp. 1603-1607.
- [76] J. M. Cioffi, "A multicarrier premier," November 1991. Also available in http://www.stanford.edu/group/cioffi/pdf/multicarrier.pdf
- [77] M. K. Simon and M.-S. Alouini, "A Simple Single Integral Representation of the Bivariate Rayleigh Distribution," *IEEE Commun. Letter*, Vol. 2, No. 5, pp. 128-130, May 1998. 2004.

Chapter 8

Variance-Bounded Rate Allocation for Constant Rate Media Streaming

²¹ In order to take advantage of time-varying nature of wireless links, transmission rate, transmission power or both rate and power are varied according to the channel quality. These adaptive transmission techniques can lead to efficient use of the communication resources, see for example [78]. As we mentioned in Chapter 1 that average transmission rate maximization algorithms for fading channels are suitable for services which have no delay constraint [79]. On the other hand fixed rate transmission with channel inversion power control for fading channel is suitable for services which have hard delay constraint. However, in order to satisfy some upper-layer QoS parameter, e.g., delay requirements, the transmission system should not only maintain an average transmission rate but also it may require to guarantee a bound on the variance of the transmission rate. For example, in case of high-quality media streaming over wireless links, there should be enough number of packets at the play-out buffer such as the video frames can be played in timely manner without any interruption (see for example [81]). In other words, in order to play out the entire video sequence smoothly, the buffer should not experience an underflow. As such the constant rate downloading from the server can avoid buffer underflow. But this constant rate downloading policy may not lead to a power efficient downloading scheme. On the other hand, from channel quality variation point of view, packet transmission should be scheduled at higher rate during relatively better channel qualities as suggested by the waterfilling algorithm. This water-filling policy, which guarantees an average rate constraint, may cause too much play-out buffer outage due to unavailability of the video frames on time. Since the video frames are to be available at the buffer before their taken out time

²¹The research work presented in this chapter is in preparation as D. V. Djonin, Md. J. Hossain, and V. K. Bhargava, "Variance-Bounded Rate Allocation for Constant Rate Media Streaming", to be submitted in *IEEE Trans. Wireless Commun.*

instants from the buffer, instead of having constant rate download some variance in the downloading rate can be tolerated.

In this chapter, we study a packet scheduling scheme for on-demand constant rate media streaming over time-varying wireless links. The objective of our proposed scheduling scheme is to minimize the average transmit power while maintaining a low receiver playout buffer outage probability. Towards this objective, we formulate the scheduling scheme as a new rate allocation scheme namely VBRA scheme. The proposed VBRA scheme, in general, is useful to any application where the OoS parameters require not only an average transmission rate guarantee but also a certain guarantee on the variance of the transmission rate. For continuous rate adaptation and channel variation case, a closed-form solution for the optimal transmission rate is obtained in terms of the channel fading gain, average rate constraint and the variance of the transmission rate. Assuming a long size media frame for constant rate media streaming, a mathematical connection between the play-out buffer outage probability and the variance of the transmission rate is established. Specifically, we propose the joint probability density function (JPDF) method based on Gaussian approximation (GA) in order to find the variance of transmission rate for a given buffer outage probability. Due to the computational complexity of JPD method, we also introduce a simpler method namely diffusion approximation (DA) method, which was applied earlier in queueing networks performance analysis (e.g., see [82]), in order to approximate the playout buffer outage probability. Using this expression, we explore the trade-offs between the average transmit power and the probability of buffer outage during the media play-out.

8.1 System Model

In order to describe the general variance bounded rate allocation problem, let us consider a transmitter that transmits information bits to a receiver as shown in Fig. 8.1. The channel is a flat fading wireless link which is time-varying in nature. We assume that the channel fading gain h is an ergodic process with PDF of $p_h(h)$. Transmission rate R and power P are adjusted according to the channel fading gain. Therefore, the transmission rate is a function of channel fading gain h and can be denoted as R(h). In continuous rate adaptation it is assumed that any positive rate is achievable i.e., R(h). In order to adjust the transmit power we assume that there exists an ideal coding scheme that allows information transmission



Figure 8.1: Scheme of the variance-bounded rate allocation.

without any transmission error. Based on the channel capacity formula the power level, P corresponding to a transmission rate of R is given by

$$\mathbf{P}(R(h),h) = \frac{\omega^2}{h} (2^{R(h)/W} - 1), \tag{8.1}$$

where W is the system bandwidth and $\omega^2 = N_0 W$ is the additive white gaussian noise power. It is important to mention that R represents the number of coded bits. If the transmission system can tolerate some specified error rate, a practical modulation and coding scheme can be used. In this case the transmit power, P can be approximated using the well known SNR gap approximation [83]

$$\mathbf{P}(R(h),h) = \frac{\omega^2 \Theta}{h} (2^{R(h)/W} - 1), \tag{8.2}$$

where Θ is a coding and modulation dependent parameter. For example, with uncoded M-QAM, Θ can be calculated as [83]

$$\Theta = \frac{1}{3} \left[Q^{-1} \left(\frac{\text{BER}_0}{4} \right) \right]^2 \tag{8.3}$$

where $Q^{-1}(\cdot)$ is the inverse standard Gaussian Q-function and BER₀ is the target BER.

For the time being let us consider the QoS parameters of the service, which has to be transmitted, are determined by two factors; one is the average transmission rate of \overline{R} [Bits/Sec] and the other is the rate variance of σ_R^2 . Later on we will see that the average transmission rate and the transmission rate variance can be determined by some different QoS parameters in case of on-demand high-quality media streaming over wireless channels.

8.2 Variance-Bounded Rate Allocation

In order to minimize the transmit power while maintaining the average rate constraint of \overline{R} and the variance of the transmission rate σ_R^2 , we can formulate the following constrained optimization problem

$$P_{av} = \min_{R(h)} E[P(R(h), h))], \qquad (8.4)$$

s.t.
$$\mathsf{E}[R(h)] \ge \bar{R}$$
 (8.5)

$$\mathsf{E}\left[(R(h) - \bar{R})^2\right] \le \sigma_R^2 \tag{8.6}$$

where the operator $E[\cdot]$ is an expectation operator taken with respect to the PDF $p_h(h)$ and P_{av} is the average transmit power for the given average rate constraint and variance bound. The inequality constraint in Eq. (8.5) is to ensure that the average transmission rate is greater or equal to the data rate, \bar{R} whereas the inequality constraint in Eq. (8.6) is to ensure that the variance of transmission rate should be kept below a given target requirement. It is important to mention that problem of maximizing average capacity of fading channel with an *average power constraint* has been studied earlier in [80]. It was shown that the power adaptation policy for fading channel which maximizes the average capacity with an average power constraint is the water-filling policy in time domain. Forming the Lagrangian of constrained optimization problem in Eqs. (8.4)-(8.6), we can write

$$J(R(h), \hat{\lambda}_1, \hat{\lambda}_2) = \mathsf{E}\left[\mathsf{P}(R(h), h)\right] - \hat{\lambda}_1(\mathsf{E}\left[R(h)\right] - \bar{R}) + \hat{\lambda}_2(\mathsf{E}\left[(R(h) - \bar{R})^2\right] - \sigma_R^2), \quad (8.7)$$

where $\hat{\lambda}_1$ and $\hat{\lambda}_2$ are the Lagrange multipliers. Taking derivative of Eq. (8.7) with respect to the variable R(h) and equating the derivative to zero, we get the necessary condition to be satisfied by the solution of optimization problem (8.4)-(8.6)

$$\frac{2^{R(h)} \times \ln 2}{h} - \hat{\lambda}_1 + 2\hat{\lambda}_2(R(h) - \bar{R}) = 0.$$
(8.8)

Eq. (8.8) is a non-linear equation in terms of R(h). The solution of Eq. (8.8) in R(h) is the optimal transmission rate $R_0(h, \hat{\lambda}_1, \hat{\lambda}_2)$ and can be obtained using MATHEMATICA as

$$R_{0}(h,\hat{\lambda}_{1},\hat{\lambda}_{2}) = \frac{1}{2\hat{\lambda}_{2}\ln 2} \left(\hat{\lambda}_{1}\ln 2 + 2\bar{R}\hat{\lambda}_{2}\ln 2 - 2\hat{\lambda}_{2} \times \text{PL} \left[\frac{(\ln 2)^{2} \times 2^{-1+\bar{R}+\frac{\hat{\lambda}_{1}}{2\hat{\lambda}_{2}}}}{h\hat{\lambda}_{2}} \right] \right)$$
(8.9)

160
where PL[z] is the "ProductLog" function that returns the principal solution for w in equation $z = w \exp(w)$. The cut-off value for the channel fading gain, h_c below which the transmission should be suspended can be obtained from Eq. (8.9) for $R_0(h, \hat{\lambda}_1, \hat{\lambda}_2) = 0$. Then h_c can be written as

$$h_c = \frac{\ln 2}{\hat{\lambda}_1 + 2\bar{R}\hat{\lambda}_2}.$$
(8.10)

Similarly taking the derivative of Eq. (8.7) with respect to $\hat{\lambda}_1$ and $\hat{\lambda}_2$ and equating them to zero, we get two equations as follows:

$$\int_{h_c}^{+\infty} R_0(h, \hat{\lambda}_1, \hat{\lambda}_2) p_h(h) dh - \bar{R} = 0, \qquad (8.11)$$

$$\int_{h_c}^{+\infty} \left[R_0(h, \hat{\lambda}_1, \hat{\lambda}_2) - \bar{R} \right]^2 p_h(h) dh - \sigma_R^2 = 0.$$
 (8.12)

Using Eq. (8.10), Eqs. (8.11) and (8.12), can be solved for $\hat{\lambda}_1$ and $\hat{\lambda}_2$ for a given \bar{R} , σ_R^2 and $p_h(h)$. Since these equations cannot be solved analytically, numerical method is used. As soon as the values of $\hat{\lambda}_1$ and $\hat{\lambda}_2$ are known, the optimal transmission rate R_0 for a given channel gain, h can be obtained from Eq. (8.9). In order to explore the shape of the dependency of the optimal transmission rate, R_0 on gain h, curves are plotted in Fig. 8.2 for different values of transmission rate variances σ_R^2 with an average rate constraint of 2[Mbits/Sec]. We assume total system bandwidth, W of 1[MHz], noise power, ω^2 of 1[mW]. We use Rayleigh fading channel model for channel fading power gain, h which leads to the exponential PDF for $p_h(h)$ i.e., [84]

$$p_h(h) = \frac{1}{h_0} \exp\left(-h/h_0\right)$$
(8.13)

where h_0 is the average channel fading power gain. We can make two observations from Fig. 8.2. First, for a given average rate constraint \overline{R} , the optimal transmission rate becomes steeper as the variance of the transmission rate decreases. This is due to the fact that lower bound on transmission rate variance does not allow differentiation of transmission rates depending on the channel quality. Second, the channel gain cut-off value h_c below which transmission is suspended, increases if a higher variation of transmission rate is allowed. This is again expected, as with the higher transmission rate variance, data can be delayed until relatively better channel quality is achieved. Theoretically there are two extreme cases. One is the infinite bound on transmission variance. In this case the optimal transmission policy approaches the water-filling policy in time domain. The other is the



Figure 8.2: Channel gain versus transmission rate ($\bar{R} = 2$ [Mbits/Sec], W = 1[MHz] and $h_0 = 10$).

zero variance-bounded case which leads to the fixed rate transmission policy. The optimal transmission rate curve is almost constant with respect to the channel gain h as shown in the Fig. 8.2 for a variance bound of 0.0001.

Using the values of $\hat{\lambda}_1$ and $\hat{\lambda}_2$, the average transmit power can be obtained as follows:

$$\mathbf{P}_{av} = \int_{h_c}^{+\infty} \frac{\omega^2 \Theta}{h} (2^{R_0(h,\hat{\lambda}_1,\hat{\lambda}_2)/W} - 1) p_h(h) dh.$$
(8.14)

The plotting of power versus variance for different average transmission rate constraints is shown in Fig. 8.3. This figure shows that as the variance of transmission rate increases, the transmit power decreases. This is obvious and due to exploiting fading diversity in time domain. As the variance of transmission rate increases, data bits can experience better channel quality and thus the required transmit power decreases. Similarly there are two extreme cases for the average transmit power. The one is the water-filling power allocation corresponding to the water-filling policy at theoretically infinite variance bound and the other is the power corresponding to the fixed rate transmission policy at zero variance bound. Obviously, the water-filling power is the lowest average transmit power for a given



Figure 8.3: Variance versus transmit power $(h_0 = 10)$.

average transmission rate.

8.3 Constant Rate Play-Out under the Receiver Buffer Outage Probability Constraint

In this section we consider the constant rate "on-demand" video play-out under the receiver buffer outage constraint as an example of the application of our proposed VBRA problem. In the sequel we will describe the system model for constant rate on-demand play-out process and define the buffer outage probability. Then we introduce two models, namely the JPDF model and the DA model in order to find the variance of the transmission rate for a given buffer outage probability during media play-out. At the end, a numerical example is presented in order to show the performance of the proposed rate allocation scheme.

8.3.1 Overall description

Let us consider a distributed mobile media streaming application where a receiving node downloads and plays out media from a power-limited transmitting node. Example of this type of media can be small size pre-stored video files on video phones. We assume that the media frames are encoded independently and each media frame has a length of F bits. The link between the two communicating nodes is time-varying. As shown in Fig. 8.1, time is slotted slot by slot and indexed as $n = 1, 2, 3 \cdots$. There are T such slots in a duration of a video frame T_f . Let the interval between two consecutive frame retrievals be $T_f[Sec]$. Each slot has the duration of T_b [Sec] and has a fixed number of channel accesses i.e., a fixed number of symbol transmissions. We assume that the fading gain h_n over a slot with index n remains roughly constant and while being identically and independently distributed from one slot to the next. In other words, we use so called block fading model [85] which is a well accepted fading model for slowly moving environment. Therefore, the transmission rate and the power remain the same over a slot but they may change from one slot to the next. In each transmission slot, according to the channel gain h_n , the transmitter chooses the transmission rate and adjusts the transmit power based on (8.2). Let the transmission rate in the *n*-th slot interval $(n = 1, 2, \dots)$ be denoted with R_n [Bits/Sec]. Since we assume the sequence of channel states h_n is i.i.d distributed, then the same can be stated for the sequence of rates R_n .

The receiver actions have two phases; the *initial download phase* and the *simultaneous* play-out and download phase. The initial download phase is always at the start of a media session and during this phase the receiver downloads bits in order to accumulate data in the play-out buffer. In this phase the media is not played out. We assume that the initial download phase has a duration of $t_0[Sec]$ which is equivalent to $I = t_0/T_b$ transmission slots. After the initial downloading period, the receiver enters into the simultaneous play-out and download phase as it starts to play-out the media frame by frame. Specifically, at the end of each T slots i.e., every $T_f = TT_b[Sec]$, a video frame of F bits is taken out of the receiver buffer and played out as shown in Fig. 8.1. Therefore, the required average transmission rate, \overline{R} is equal to the number of bits per frame divided by the frame duration i.e.,

$$\bar{R} = \frac{F}{TT_b} [\text{Bits/Sec}]. \tag{8.15}$$

At the end of a frame duration of T_f [Sec], if the play-out frame is not available at the buffer, the receiver declares a play-out outage and starts the session from the beginning. The viewer wants to view the entire media content played out as smoothly as possible with minimal freezes and outages.

For notational convenience, we use different time index k to denote the k-th frame take out time as shown in Fig. 8.1. The number of bits downloaded during the k-th frame is

$$Y_k = (R_{kT+1} + R_{kT+2} + \dots + R_{kT+T}) T_b$$
(8.16)

for k = 1, 2, ... If the number of slots per frame duration, T is sufficiently large, according to the central limit theorem distribution of Y_k approaches Gaussian for i.i.d distributed sequence of rates R_i . Since no frame is taken out of the buffer during the initial download phase of I slots and a frame of F bits is taken out every $T_f[Sec]$ after that, the receiver buffer occupancy $X_k[Bits]$ at the end of frame index k = 1, 2, 3, ... can be expressed as

$$X_{k} = \begin{cases} \sum_{i=1}^{k} Y_{i} & \text{if } k \leq I\\ \sum_{i=1}^{I} Y_{i} + \sum_{i=I+1}^{k} (Y_{i} - F) & \text{if } k > I \end{cases}.$$
(8.17)

For sufficiently long frame length T, random variable X_k also follows the Gaussian distribution.

The design goal we analyze is the minimization of the probability of receiver buffer outage during the play-out phase. The receiver buffer outage occurs when the receiver buffer occupancy, X_k is less than zero in any frame index k (k > I). In view of the variancebounded rate allocation proposed in Section 8.2 and the receiver buffer model described above, we next demonstrate how to find the outage probability for a given variance of the transmission rate.

8.3.2 Joint PDF methodology for calculation of the receiver buffer outage probability

The probability of receiver buffer outage during the transmission of a file of length Z frames for buffer dynamics given in (8.17) can be analytically expressed as

$$P_o = 1 - \text{Prob}[X_{I+1} > 0, X_{I+2} > 0, \dots, X_Z > 0].$$
 (8.18)

This can be further expressed in terms of the joint PDF $p(X_{I+1}, X_{I+2}, ..., X_Z)$ of random variables $X_{I+1}, X_{I+2}, ..., X_Z$ as

$$P_o = 1 - \int_0^\infty \cdots \int_0^\infty p(X_{I+1}, X_{I+2}, \dots, X_Z) dX_{I+1} dX_{I+2} \cdots dX_Z.$$
(8.19)

Now, based on Eq. (8.17) it can be concluded that the sequence of receiver buffer occupancies constitutes a Markov chain, i.e. that the joint PDF $p(X_{I+1}, X_{I+2}, ..., X_Z)$ can be expressed as

$$p(X_{I+1}, X_{I+2}, \dots, X_Z) = p(X_{I+1})p(X_{I+2}|X_{I+1})\cdots p(X_Z|X_{Z-1}).$$
(8.20)

As we have discussed previously, for a sufficiently large frame length, the buffer occupancy follows a Gaussian distribution. Furthermore, conditional distributions $p(X_k|X_{k-1})$, for $I < k \leq Z$ follow the Gaussian distribution as well. Namely,

$$p(X_{I+1}) = \mathcal{N}_{X_{I+1}}(TIT_b\bar{R}, (I+1)T\sigma_R^2)$$
(8.21)

$$p(X_k | X_{k-1}) = \mathcal{N}_{X_k}(X_{k-1}, T\sigma_R^2), I < k \le Z$$
(8.22)

where $\mathcal{N}_Z(\bar{Z}, \sigma_R^2)$ is a Gaussian distribution of random variable Z with expectation \bar{Z} and variance σ_R^2 . Substituting (8.21) into (8.20) and then into (8.19), we can get the expression for the outage probability given in terms of parameters $I, T, \sigma_R^2, \bar{R}$ as

$$P_{o} = 1 - \int_{0}^{\infty} \cdots \int_{0}^{\infty} \mathcal{N}_{X_{I+1}}(ITT_{b}\bar{R}, (I+1)T\sigma_{R}^{2}) \times \mathcal{N}_{X_{I+2}}(X_{I+1}, T\sigma_{R}^{2}) \times \cdots \times \mathcal{N}_{X_{Z}}(X_{Z-1}, T\sigma_{R}^{2})dX_{I+1} \cdots dX_{Z}.$$
(8.23)

For sufficiently large T this approach will give the precise values for the probability of outage in the receiver buffer. However, from a computational point of view, using Eq. (8.23) is equivalent to computing the cumulative distribution function of a multivariate Gaussian distribution that is computationally intensive for large dimensional distributions.

8.3.3 Diffusion approximation for calculation of the receiver buffer outage probability

When the number of frames Z becomes large, the calculation of receiver buffer outage probability using the JPDF methodology is practically impossible due to its computational complexity. In this section we, therefore, introduce DA model which offers an upper bound

and relatively simple approximative evaluation of the buffer outage probability. Since sequence of transmission rates $\{R_n\}$ are independent random variables, the play-out buffer occupancy at time $T_f k$, X_k can be modelled as a random walk (RW). The simplest definition of a RW is a sum of a number of i.i.d. random variables [86]. When the number of frames Z becomes very large, a more physical approach is applicable and based on passage from discrete RW to a diffusion process that requires the following assumptions [86]

- The slot length approaches zero, i.e. T_b → 0, and the time axis is described with continuous variable t. Keep the frame length of T_f[Sec] fixed. The average increments *R*T_b in buffer occupancy during a slot approach zero, and the buffer occupancy is also represented with a continuous variable.
- The data retrieval process from the buffer is performed evenly throughout the frame of length $T = T_f/T_b \rightarrow \infty$ slots. Therefore $F/T \rightarrow 0$ bits is retrieved from the buffer in each slot.

Let the PDF of buffer occupancy, x at time t, be $b(x, t)^{22}$. Then according to the DA, the dynamics of the buffer can be expressed by to the diffusion equation as follows: [86]

$$\frac{\delta b}{\delta t} = D_{\text{diff}} \frac{\delta^2 b}{\delta x^2} - v_{\text{diff}} \frac{\delta b}{\delta x}$$
(8.24)

where the coefficient D_{diff} is known as the diffusion constant and given as

$$D_{\rm diff} = \frac{\sigma_R^2}{2T_b} \tag{8.25}$$

and v_{diff} is the average rate of diffusion. The detailed derivation of diffusion equation can be found in [86]. The solution of the diffusion equation subject to the initial condition that the buffer is at x' with probability 1 at time t' i.e., $b(x', t') = \delta(x - x')$ can be written as [86]

$$b(x,t|x',t') = \frac{1}{\sqrt{2\pi(\sigma_R^2/T_b)(t-t')}} \times \exp\left[-\frac{(x-x'-v_{\text{diff}}(t-t'))}{(2\sigma_R^2/T_b)(t-t')}\right].$$
 (8.26)

It is important to mention that the solution is Gaussian at all time times which can also be shown from central limit theorem.

 $^{^{22}}$ In order to avoid notational confusion, we denote the continuous random variable buffer occupancy and the time as x and t, respectively.

According to our transmission framework and buffer dynamics equation, if the buffer occupancy falls below zero, there is an outage and the session starts from the beginning. So, the buffer occupancy process X which is a diffusion process has an absorbing point, at x = 0 [86]. An absorbing boundary is a surface with the property that it removes from the system any random walker that comes into contact with it. Therefore, the diffusion Eq. (8.24) is to solved subject to the boundary condition $b(0, t_0) = 0$. Using this boundary condition the solution can be written as [86]

$$b(x,t|x_0,t_0) = b_F(x,t|x_0,t_0) - b_F(x,t|-x_0,t_0)$$
(8.27)

where $b_F(x, t|x_0, t_0)$ is the PDF in the absence of boundaries for the buffer occupancy at time t of the buffer which is at x_0 at time t_0 :

$$b_F(x,t|x_0,t_0) = \frac{1}{\sqrt{2\pi(\sigma_R^2/T_b)(t-t_0)}} \times \exp\left[-\frac{(x-x_0)^2}{2(\sigma_R^2/T_b)(t-t_0)}\right].$$
(8.28)

Note that based on our second assumption, rate of diffusion in the play-out phase is $v_{\text{diff}} = 0$, as the average incoming traffic in the buffer is completely offset by the retrieval process. The survival probability of the buffer or the probability that its occupancy does not goes below zero by time t, is denoted by $S(t|x_0, t_0)$ and is given by

$$S(t|x_0, t_0) = \int_0^\infty b(x, t|x_0, t_0) dx.$$
(8.29)

Using Eqs. (8.27) and (8.28), Eq. (8.29) can be written as

$$S(t|x_0, t_0) = \operatorname{erf}\left(\frac{x_0}{\sqrt{(2\sigma_R^2/T_b)(t - t_0)}}\right)$$
(8.30)

where $erf(\cdot)$ is the "error function" and defined as

$$\operatorname{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z \exp(-u^2) \, du.$$
 (8.31)

It is important to mention that the survival probability is conditional probability conditioned on that the buffer occupancy is x_0 at time t_0 . Since no frames are taken out of the buffer during the initial download period t_0 , the buffer occupancy at the end of time t_0 can be approximated as free-space (free space means without any boundaries) diffusion equation solution with $v_{\text{diff}} = \bar{R}$, as follows: [86]

$$b_0(x_0, t_0) = \frac{1}{\sqrt{2\pi(\sigma_R^2/T_b)t_0}} \times \exp\left[-\frac{(x_0 - \bar{R}t_0)^2}{2(\sigma_R^2/T_b)t_0}\right].$$
(8.32)

Then the probability that the buffer will have no outage during play-out period of $t = ZT_f[Sec]$ with an initial download period of $t_0[Sec]$, S_{avg} can be obtained by averaging the conditional survival probability, $S(t|x_0, t_0)$ over the PDF of x_0 , $b_0(x_0, t_0)$ and can be expressed as

$$S_{\text{avg}}(t|t_0) = \int_0^{+\infty} S(t|x_0, t_0) b_0(x_0, t_0) dx_0.$$
(8.33)

Note the above equation can only be used for sufficiently large initial download period. Using Eqs. (8.30) and (8.32) in Eq. (8.33), finally the outage probability with an initial download phase of t_0 [Sec], $P_o^{DA}(t|t_0)$ can be expressed as

$$P_o^{\text{DA}}(t|t_0) = 1 - \sqrt{\frac{T_b}{2\pi\sigma_R^2}} \int_0^{+\infty} \exp\left[-\frac{(x_0 - \bar{R}t_0)^2}{2(\sigma_R^2/T_b)t_0}\right] \times \operatorname{erf}\left(\frac{x_0}{\sqrt{2(\sigma_R^2/T_b)(t - t_0)}}\right) dx_0.$$
(8.34)

From Eq. (8.34) it is possible to find the variance of transmission rate for a given outage probability. Then the value of variance can be used to find the optimal transmission rate as described in Section 8.2.

In order to compare two different methods described above, we plot outage probability versus variance in Fig. 8.4. For this figure we assume that a frame length of 50 slots and frame rate of 25 frames per second (fps) is used. Therefore, the frame duration is $T_f = 1/25$ [Sec]. We assume an initial download phase of 1 slot, while the play-out phase (due to computational complexity associated with Eq. 8.23) is chosen to be only 2 slots. Rather low average transmission rate of $\overline{R} = 1$ [kbits/sec] is chosen only as an illustration and to ensure that outage probabilities are on the order of $10^{-6} - 10^{-1}$.

This figure shows that DA model offers slightly higher outage probability than the JPDF method for a given variance bound. This is due to the fact that the DA model assumes that data retrieval process is performed evenly throughout the frame length instead of taking out frame at the end of a frame duration. So, the outage probability provided by the DA is more stringent and the DA model offers the upper bound on outage probability. In practice, the buffer will experience lower outage probability than the calculated one from DA model.

8.3.4 Results

In this section we show some numerical examples in order to show the performance of our proposed scheme. Specifically we plot the optimal trade-offs between the transmit



Figure 8.4: Buffer outage probability versus transmission rate variance (W = 1[MHz], and $h_0 = 10$).



Figure 8.5: Buffer outage probability versus transmit power (W = 1[MHz], and $h_0 = 10$).

power and the buffer outage probability using DA. In our setting we have used a video length of 10000 frames. Therefore total duration of the video is 400[Sec]. We assume an initial download phase of 0.2[Sec]. The optimal power-outage trade-offs plotted in Fig. 8.5 shows that as the outage probability increases the required transmit power decreases. This is obvious as the higher outage probability allows to have higher transmission rate variance and consequently more fading diversity is obtained. It is also remarkable that allowing a very low buffer outage can save lot more transmit power.

8.4 Conclusions

In this chapter we have proposed and studied a VBRA scheme which is useful for delay constraint communications e.g., streaming media play-out over time-varying wireless channels. In our proposed transmission framework the transmission rate is adjusted according to the wireless channel quality. As such, the transmit power is minimized and an average rate constraint is satisfied. Also, the proposed rate-allocation guarantees a transmission rate variance constraint that in turn controls the buffer underflow during play-out. Our transmission framework is optimal if the frame length is sufficiently large for Gaussian approximation to hold.

Bibliography

- [78] A. J. Goldsmith and S.-G. Chua "Variable-rate variable-power MQAM for fading channels," *IEEE Trans. Commun.*, vol. 45, pp.1218-1230, Oct. 1997.
- [79] D. N. C. Tse and S. V. Hanly, "Multiple access fading channels-Part I: Polymatriod structure, optimal resource allocation and throughput capacities", *IEEE Trans. Inform. Theory*, vol. 44 no. 7, pp. 2796-2815, Nov. 1998.
- [80] A. J. Goldsmith and P. P. Varaiya "Capacity of Fading Channels with Channel Side Information," *IEEE Trans. Inform. Theory*, vol. 43, pp.1986-1992, Nov. 1997.
- [81] Y. Li, A. Markopoulou, N. Bambos, and A. J. Apostolopoulos, "Joint Power/Playout Control Schemes for Media Streaming over Wireless Links," in *Proc. IEEE Packet Video 2004*, Irvine, CA, Dec. 2004. Also available online http://www.stanford.edu/ amarko/PAPERS/pv04.pdf.
- [82] H. Kobayashi, "Application of the diffusion to queueing networks I: equilibrium queue distribution," *Jour. Ass. Computing Mach.*, vol. 21, pp. 316-328, Apr. 1974.
- [83] J. M. Cioffi, "A multicarrier premier," Nov. 1991. Also available in http://www.stanford.edu/group/cioffi/pdf/multicarrier.pdf.
- [84] M. K. Simon and M.-S. Alouini, "Digital Communications over Fading Channels," 2nd Edition, New Jersey : John Wiley & Sons, 2005.
- [85] E. Biglieri, G. Caire, and G. Taricco, "Limiting performance of block-fading channels with multiple antennas," *IEEE Trans. Inform. Theory*, vol. 47, pp. 1273-1289, May 2001.
- [86] G. H. Weiss, "Aspects and Applications of the Random Walk," North-Holland, Elsevier Science, 1994.

Chapter 9

Conclusion

In this dissertation, our contribution is the design of innovative scheduling schemes via a *cross-layer design approach* that jointly considers the time-varying nature of wireless channels as well as the required QoS parameters of different services. Specifically, this dissertation makes the following four major contributions.

First, we have developed modulation-assisted two-user opportunistic scheduling schemes. Specifically, in Chapter 2, we have proposed a power-controlled hierarchical modulationassisted TBS. The information access probability and some radio link layer performance of the proposed power-controlled TBS scheme have been analyzed in this chapter. Presented numerical results show that the power-controlled hierarchical modulation-assisted TBS provides lower average buffer occupancy (i.e., queuing delay) and loss probability for the packets at the transmission buffer, at the expense of transmit power, compared with the classical SBS scheme. The required additional power decreases as the number of users increases in the system. We have compared our hierarchical scheme with a uniform constellation-based mini-slotted scheme. This mini-slotted scheme provides similar higher link layer performance as the hierarchical scheme but requires higher transmit power. Since the power-controlled modulation-assisted TBS includes the second best user in each transmission, it requires a certain amount of power compared with the classical SBS scheme. Therefore in Chapter 3, we have proposed a rate adaptive hierarchical modulationassisted TBS scheme. The proposed rate adaptive modulation-assisted TBS scheme does not include the second best user in each transmission. Instead, it proposes to transmit information to the second best user opportunistically by joint consideration of the channel qualities of the users. We have evaluated the information access probability and system SE in closed-form for an i.i.d fading environment in Chapter 3. Numerical results presented in this chapter show that the rate adaptive modulation-assisted TBS scheme almost doubles the frequency of information access for users without any degradation of overall SE, compared with the classical SBS. In Chapter 4, we have derived expressions for the SE of

the users for an i.n.d. fading environment. Using this expression, we have compares the fairness of the rate adaptive TBS scheme with that of the classical single user scheduling schemes. From a numerical example, we have observed that the rate adaptive TBS scheme increases the system's fairness without any degradation of system SE, compared with the classical SBS. In order to further improve system fairness we have proposed a rate adaptive HTS. The novelty of the HTS is that it significantly improves the system's fairness without any degradation of system SE compared with the SBS scheme; it cannot, however, achieve system fairness as good as the equivalent PFS, which suffers from a certain degradation of SE. In Chapter 5, we have analyzed some link layer performances; delay statistics and buffer distribution of the proposed rate adaptive TBS and HTS schemes via queuing analysis. Presented numerical results in this chapter show that TBS can significantly reduce the queuing delay of the packets at the transmission buffer. Interestingly, this improved delay performance is gained without any SE cost. For an i.n.d. fading environment, the HTS scheme significantly improves the delay performance of the far-end user, compared with that of the TBS and SBS schemes. It also achieves a delay performance that is close to that of the equivalent PFS (or NSS) scheme for the far-end user. However, the NSS scheme significantly increases the queuing delay of the packets for the near-end user.

Second, we have developed, in Chapter 6, an adaptive scheduling protocol for multiresolution image/video data transmission over time-varying channels. The proposed scheme not only takes advantage of hierarchical constellations but also uses an adaptive scheduling protocol to give different transmission priorities to different resolution levels. A Markov chain model of this protocol has led to closed-form expressions for the average packet loss rate and the average transmission rate. Numerical results show that the proposed scheme can control the relative transmission priorities in terms of transmission rate as well as loss rate of different resolution levels by varying the asymmetry of the constellation and the maximum number of allowed retransmissions. In Chapter 7, we have also proposed an integrated packet scheduling and unequal error protection strategy for real time packetized multi-resolution (layered) image/video data transmission over correlated channels. In order to the select transmit power level, a constellation priority parameter and the packets from different resolution levels, the proposed system considers the timeliness requirements and the importance levels of different packets as well as time-varying channel property. A joint optimization of the proposed transmission framework has been accomplished by formulating it as a MDP. The proposed hierarchical scheme has also been compared with an optimized scheme that uses uniform signal constellation. The latter scheme requires higher average transmit power than the hierarchical scheme in order to maintain the same average distortion of the received image/video data.

Third, in Chapter 7 we have developed, link adaptation algorithms for a multicarrier system for generalized delay constrained communications over time-varying channels. Specifically, we have proposed two different transmission frameworks for delayconstrained OFDM communication. One is the bit-loading framework, which minimizes the aggregate BER of a constant power OFDM system while meeting a delay constraint of the packets at the transmission buffer. The other is the bit- and power- loading algorithm, which seeks to minimize the total average transmit power while maintaining a target delay constraint. Optimal loading policies in both cases are obtained using LP methodology formulating them as the MDP. Since finding the optimal policies and implementing them are highly complex in reality, if not impossible, we also explore a suboptimal loading algorithm based on an equivalent single carrier optimal loading scheme and a greedy approach. Since the suboptimal loading algorithm requires only finding and storing the optimal policy of the equivalent single carrier case with a fewer number of states, it is simple to realize it practically. The selected numerical results show that the suboptimal loading scheme in both cases attains a tradeoff performance very close to the optimal one. In Chapter 7, we have also explored the effect of correlation between the subcarriers' fading gains of a MCM system on its bit- and power-loading in delay-constrained and no delay-constrained communications. For the no delay-constrained case, the correlation between subcarrier fading gains does not have any influence on the loading policy. Therefore, the average total transmit power is not influenced by the correlation between the subcarriers' fading gains. For discrete rate adaptation as well as discrete channel gain variation, the optimal loading policies for both cases are obtained through a greedy allocation. Assuming discrete rate adaptation and discrete channel gain variation, numerical examples have been demonstrated for different degrees of correlation. The results show that the required transmit power in the hard delay-constrained case increases as the correlation increases. However, the degree of correlation does not have any significant effect on the required transmit power in the no delay-constrained case. Although we have considered only correlation among the carriers' fading gains, it is worthy to mention that time correlation does not influence our conclusions.

Fourth, in Chapter 8 we have devised a novel VBRA scheme for delay-constrained communications, e.g., media streaming over time-varying channels. In this chapter, we studied a VBRA problem that is useful for streaming media play-out over time-varying wireless channels. In our proposed transmission framework, the transmission rate is adjusted according to the wireless channel quality; the transmit power is minimized and an average rate constraint is satisfied. Also, the proposed rate-allocation guarantees a transmission rate variance constraint that in turn controls the buffer underflow during play-out. Our transmission framework is optimal if the frame length is sufficiently large for Gaussian approximation to hold.

9.1 Future Research Challenges

Our contributions in this dissertation, have not only provided innovative scheduling algorithms for time-varying wireless links, but also opened up exciting research avenues for future investigation. These research challenges are described below.

9.1.1 Opportunistic subcarrier allocation for OFDM-based system

In our proposed TBS schemes, we considered a single-carrier system. However, next generation, wireless networks, e.g., WiMAX, will use OFDMA as a multiple access technique. With hierarchical modulation, one subcarrier of an OFDMA system can be shared among more than one user. Therefore, it is interesting to design and analyze the performance of an OFDMA-based system with opportunistic subcarrier allocations in conjunction with hierarchical transmission. We conjecture that since the hierarchical transmission scheme allows an OFDM subcarrier to be shared among more than one user it can achieve better system performance.

9.1.2 Designing multiple-best user scheduling scheme

In our thesis, we have considered multiple best user scheduling scheme using power controlled hierarchical constellations. As more users are included in the system, required average transmit power increases. The rate adaptive modulation assisted TBS scheme does not degrade overall SE but improves queuing delay performance. Further delay improvement can be obtained by including third best user in the system cleverly. However, designing such scheme with rate adaptive hierarchical modulations requires complicated design. Specifically, three dimensional joint PDF of CNRs of the first, second and best users needs to be considered. Further investigation in this direction can be interesting but challenging.

9.1.3 Rate adaptation for hybrid ARQ-based system for real-time multi-resolution data transmission

Although we have used hierarchical modulation for unequal error protection for multiresolution image/video data transmission over time-varying channels, our transmission frameworks in this regard can be generalized for any unequal error protection system. Specifically, it is interesting to design an adaptive and optimized scheme for real-time multi-resolution with a hybrid ARQ scheme. In this case, the scheme should change the coding rate for the current packets based on the timeliness of the packets from different resolution levels and the dynamics of time-varying channels.

9.1.4 Hybrid policy for media streaming over fading channels

Our proposed VBRA allocation schemes for media streaming over time varying channels is optimal among a set of policies in which there is only knowledge of the channel information and no knowledge of the receiver buffer occupancy. Finding a policy with buffer knowledge is indeed a non-stationary policy. Implementation of a non-stationary policy is restricted by its complexity. Developing a hybrid policy based on our VBRA scheme and the partial receiver's buffer knowledge is challenging. Obviously, without full knowledge of the buffer, this policy would be suboptimal. However, having partial buffer knowledge can further improve system performance.

9.1.5 Effect of imperfect channel state and slow adaptation for time channels

All of the adaptive schemes proposed in this dissertation assume that the channel information is perfectly known at the transmitter. However, in practice the channel state information may be outdated or only known partially, e.g., in the form of acknowledgment signals for the ARQ system. Studying the performance of our proposed scheme with an imperfect channel state is another exciting future research direction. It is obvious that the performance of our proposed scheme will degrade with an imperfect channel state at the transmitter. Based on such future work, robust adaptive schemes can be designed in these contexts.

Channel adaptive schemes also require a great amount of overhead, which in turn reduces the effective data transmission rate, especially for multiuser, multi-antenna, and multicarrier transmission systems. Therefore, researchers are attempting to develop adaptive schemes with limited feedback. Examples of these recently developed schemes include blind adaptive modulation schemes and binary feedback-based adaptive modulation schemes. Therefore, it is very important to design adaptive transmission schemes that adapt the transmission parameters according to channel statistics, e.g., average channel quality or amount of fading, instead of instantaneous channel quality. In future research, it would be very practical to extend our adaptive schemes for the more general multiuser, OFDM and MIMO systems.

Appendices

Appendix A

In this appendix, we derived the integral in Eq. (3.13). Using the expression of the joint PDF in Eq. (3.3), the integral $\int_{\gamma_{r,s}}^{\gamma_{r,s+1}^2} \int_{\gamma_{r,s}}^{g_{(1)}} p_{g_{(1)}g_{(2)}}(g_{(1)}, g_{(2)}) dg_{(2)}dg_{(1)}$ in Eq. (3.13) can be evaluated as follows:

$$\begin{split} &\int_{\gamma_{r,s}^{2}}^{\gamma_{r,s+1}^{2}} \int_{\gamma_{r,s}^{1}}^{g_{(1)}} p_{g_{(1)}g_{(2)}}(g_{(1)},g_{(2)})dg_{(2)}dg_{(1)} \\ &= \int_{\gamma_{r,s}^{2}}^{\gamma_{r,s+1}^{2}} \int_{\gamma_{r,s}^{1}}^{g_{(1)}} \frac{K(K-1)}{g_{0}^{2}} \exp\left(-g_{(1)}/g_{0}\right) \exp\left(-g_{(2)}/g_{0}\right) \\ &\times \left[1 - \exp\left(-g_{(2)}/g_{0}\right)\right]^{K-2} dg_{(2)}dg_{(1)} \\ &= \int_{\gamma_{r,s}^{2}}^{\gamma_{r,s+1}^{2}} \int_{\gamma_{r,s}^{1}}^{g_{(1)}} \frac{K(K-1)}{g_{0}^{2}} \sum_{k=0}^{K-2} \binom{K-2}{k} (-1)^{k} \exp\left(-g_{(1)}/g_{0}\right) \\ &\times \exp\left(-(k+1)g_{(2)}/g_{0}\right) dg_{(2)}dg_{(1)} \\ &= \frac{K(K-1)}{g_{0}^{2}} \sum_{k=0}^{K-2} \binom{K-2}{k} (-1)^{k} \int_{\gamma_{r,s}^{2}}^{\gamma_{r,s+1}^{2}} \exp\left(-g_{(1)}/g_{0}\right) \\ &\times \left[\int_{\gamma_{r,s}^{1}}^{g_{(1)}} \exp\left(-(k+1)g_{(2)}/g_{0}\right) dg_{(2)}\right] dg_{(1)} \\ &= \sum_{k=0}^{K-2} \frac{K(K-1)}{k+1} \binom{K-2}{k} (-1)^{k} \left[\exp\left(-(k+1)\gamma_{r,s}^{1}/g_{0}\right) \\ &\times \left[\exp\left(-\gamma_{r,s}^{2}/g_{0}\right) - \exp\left(-\gamma_{r,s+1}^{2}/g_{0}\right)\right] \\ &- \frac{\exp\left(-(k+2)\gamma_{r,s}^{2}/g_{0}\right) - \exp\left(-(k+2)\gamma_{r,s+1}^{2}/g_{0}\right)}{k+2} \right]. \end{split}$$

$$(9.1)$$

Similarly using the expression of the joint PDF in Eq. (3.3), the integral

$$\int_{\gamma_{r,s}^{2}}^{\gamma_{r,s+1}^{2}} \int_{\gamma_{r,s}^{1}}^{\gamma_{r,s}^{2}} p_{g_{(1)}g_{(2)}}(g_{(1)}, g_{(2)}) dg_{(2)} dg_{(1)} \text{ in Eq. (3.18) can be evaluated as:} \\
= \int_{\gamma_{r,s}^{2}}^{\gamma_{r,s+1}^{2}} \int_{\gamma_{r,s}^{1}}^{\gamma_{r,s}^{2}} p_{g_{(1)}g_{(2)}}(g_{(1)}, g_{(2)}) dg_{(2)} dg_{(1)}, \gamma_{r,s}^{1} \leq \gamma_{r,s}^{2} \\
= \int_{\gamma_{r,s}^{2}}^{\gamma_{r,s+1}^{2}} \int_{\gamma_{r,s}^{1}}^{\gamma_{r,s}^{2}} \frac{K(K-1)}{(g_{0})^{2}} \exp\left(-g_{(1)}/g_{0}\right) \exp\left(-g_{(2)}/g_{0}\right) \left[1 - \exp\left(-g_{(2)}/g_{0}\right)\right]^{K-2} \\
= \int_{\gamma_{r,s}^{2}}^{\gamma_{r,s+1}^{2}} \int_{\gamma_{r,s}^{1}}^{\gamma_{r,s}^{2}} \frac{K(K-1)}{(g_{0})^{2}} \sum_{k=0}^{K-2} \binom{K-2}{k} (-1)^{k} \exp\left(-g_{(1)}/g_{0}\right) \\
\times \exp\left(-(k+1)g_{(2)}/g_{0}\right) dg_{(1)} dg_{(2)}$$
(9.2)

Switching the order of the summation and the integrals, we can write Eq.

$$\int_{\gamma_{r,s}^{2}}^{\gamma_{r,s+1}^{2}} \int_{\gamma_{r,s}^{1}}^{\gamma_{r,s}^{2}} p_{g_{(1)}g_{(2)}}(g_{(1)},g_{(2)}) dg_{(2)} dg_{(1)}, \gamma_{r,s}^{1} \leq \gamma_{r,s}^{2}$$

$$= \frac{K(K-1)}{(g_{0})^{2}} \sum_{k=0}^{K-2} {\binom{K-2}{k}} (-1)^{k} \int_{\gamma_{r,s}^{2}}^{\gamma_{r,s+1}^{2}} \exp\left(-g_{(1)}/g_{0}\right) dg_{(1)}$$

$$\times \int_{\gamma_{r,s}^{1}}^{\gamma_{r,s}^{2}} \exp\left(-(k+1)g_{2}/g_{0}\right) dg_{2}$$

$$= \left[\exp\left(-\gamma_{r,s}^{2}/g_{0}\right) - \exp\left(-\gamma_{r,s+1}^{2}/g_{0}\right)\right] \times \sum_{k=0}^{K-2} \frac{K(K-1)}{k+1} {\binom{K-2}{k}}$$

$$\times (-1)^{k} \left[\exp\left(-(k+1)\gamma_{r,s}^{1}/g_{0}\right) - \exp\left(-(k+1)\gamma_{r,s}^{2}/g_{0}\right)\right]. \quad (9.3)$$

Appendix B

Integral:
$$I_1^j(t_1, t_2) = \int_{t_1}^{t_2} p_{g_j}(x) \int_0^x p_{g_{(1)}^*}(z) dz dx.$$
 (9.4)

For Rayleigh fading channel using Eqs. (4.1), (4.2) and (4.7), the integral in Eq. (9.4) can be evaluated as follows: [43]

$$\begin{split} I_{1}^{j}(t_{1},t_{2}) &= \int_{t_{1}}^{t_{2}} \frac{1}{\bar{g}_{j}} \exp\left(-\frac{x}{\bar{g}_{j}}\right) \int_{0}^{x} \sum_{k=1, k \neq j}^{K} p_{g_{k}}(z) \prod_{m=1, m \neq k, j}^{K} P_{g_{m}}(z) dx \\ &= \int_{t_{1}}^{t_{2}} \frac{1}{\bar{g}_{j}} \exp\left(-\frac{x}{\bar{g}_{j}}\right) \prod_{n=1, n \neq j}^{K} \left(1 - \exp\left(-\frac{x}{\bar{g}_{n}}\right)\right) dx \\ &= \int_{t_{1}}^{t_{2}} \frac{1}{\bar{g}_{j}} \exp\left(-\frac{x}{\bar{g}_{j}}\right) \sum_{\tau' \in T_{j}^{K}} \operatorname{sign}(\tau') \exp\left(-\tau'x\right) dx \\ &= \int_{t_{1}}^{t_{2}} \sum_{\tau' \in T_{j}^{K}} \frac{\operatorname{sign}(\tau')}{\bar{g}_{j}} \exp\left(-x(\tau' + \frac{1}{\bar{g}_{j}})\right) dx \\ &= \sum_{\tau' \in T_{j}^{K}} \frac{\operatorname{sign}(\tau')}{(\tau'\bar{g}_{j} + 1)} \left[\exp\left(-\frac{t_{1}(\tau'\bar{g}_{j} + 1)}{\bar{g}_{j}}\right) - \exp\left(-\frac{t_{2}(\tau'\bar{g}_{j} + 1)}{\bar{g}_{j}}\right) \right] (9.5) \end{split}$$

where the notations τ' has been adopted from the [43] and the references there in. More specifically, T_j^K is the set obtained by expanding the product $\prod_{n=1,n\neq j}^K \left(1 - \exp\left(-\frac{y}{\bar{g}_n}\right)\right)$ then by taking the natural logarithm of each term, and $\operatorname{sign}(\tau)$ is the corresponding sign of each term in the expansion.

Integral:
$$I_2^j(t_1, t_2, t_3) = \int_{t_1}^{t_2} p_{g_j}(x) \int_0^{t_3} p_{g_{(1)}^*}(z) dz dx.$$
 (9.6)

Similarly for Rayleigh fading channel using Eqs. (4.1), (4.2) and (4.7) in Eq. (9.6), the integral can be evaluated in closed-form as follows:

$$\begin{split} I_{2}^{j}(t_{1}, t_{2}, t_{3}) &= \int_{t_{1}}^{t_{2}} \frac{1}{\bar{g}_{j}} \exp\left(-\frac{x}{\bar{g}_{j}}\right) \int_{0}^{t_{3}} p_{g_{(1)}^{*}}(z) dz dx \\ &= \left(\int_{t_{1}}^{t_{2}} \frac{1}{\bar{g}_{j}} \exp\left(-\frac{x}{\bar{g}_{j}}\right) dx\right) \left(\int_{0}^{t_{3}} \sum_{k=1, k \neq j}^{K} p_{g_{k}}(z) \prod_{l=1, l \neq k, j}^{K} P_{g_{l}}(z) dz\right) \\ &= \left(\exp\left(-\frac{t_{1}}{\bar{g}_{j}}\right) - \exp\left(-\frac{t_{2}}{\bar{g}_{j}}\right)\right) \left(\prod_{k=1, k \neq j}^{K} P_{g_{k}}(t_{3})\right) \\ &= \left(\exp\left(-\frac{t_{1}}{\bar{g}_{j}}\right) - \exp\left(-\frac{t_{2}}{\bar{g}_{j}}\right)\right) \left(\prod_{k=1, k \neq j}^{K} \left(1 - \exp\left(-\frac{t_{3}}{\bar{g}_{k}}\right)\right)\right). (9.7) \end{split}$$

Integral:
$$I_3^j(t_1, t_2, t_3) = \int_{t_1}^{t_3} p_{g_j}(x) \int_{t_3}^x p_{g_{(1)}^*}(z) dz dx.$$
 (9.8)

Similarly, the integral in Eq. (9.8) can be evaluated as follows:

$$\begin{split} I_{3}^{j}(t_{1}, t_{2}, t_{3}) &= \int_{t_{1}}^{t_{2}} p_{g_{j}}(x) \int_{t_{3}}^{x} p_{g_{(1)}^{*}}(z) dz dx \\ &= \int_{t_{1}}^{t_{2}} p_{g_{j}}(x) \left(\int_{0}^{x} p_{g_{(1)}^{*}}(z) dz - \int_{0}^{t_{3}} p_{g_{(1)}^{*}}(z) dz \right) dx \\ &= \int_{t_{1}}^{t_{2}} p_{g_{j}}(x) \left(\prod_{k=1, k \neq j}^{K} \left(1 - \exp\left(-\frac{x}{\bar{g}_{k}} \right) \right) - \left(1 - \exp\left(-\frac{t_{3}}{\bar{g}_{k}} \right) \right) \right) dx \\ &= \int_{t_{1}}^{t_{2}} \frac{1}{\bar{g}_{j}} \exp(-\frac{x}{\bar{g}_{j}}) \times \sum_{\tau \in T_{j}^{K}} \operatorname{sign}(\tau) \left[\exp\left(-\tau x \right) - \exp\left(-\tau t_{3} \right) \right] dx \\ &= \sum_{\tau \in T_{j}^{K}} \frac{\operatorname{sign}(\tau)}{\bar{g}_{j}} \int_{t_{1}}^{t_{2}} \left[\exp\left(-\frac{(\tau \bar{g}_{j} + 1)x}{\bar{g}_{j}} \right) - \exp\left(-\frac{x}{\bar{g}_{j}} \right) \exp\left(-\tau t_{3} \right) \right] dx \\ &= \sum_{\tau \in T_{j}^{K}} \operatorname{sign}(\tau) \left[\frac{1}{(\tau \bar{g}_{j} + 1)} \left(\exp\left(-\frac{(\tau \bar{g}_{j} + 1)t_{1}}{\bar{g}_{j}} \right) - \exp\left(-\frac{(\tau \bar{g}_{j} + 1)t_{2}}{\bar{g}_{j}} \right) \right) \\ &- \left(\exp\left(-\frac{t_{1}}{\bar{g}_{j}} \right) - \exp\left(-\frac{t_{2}}{\bar{g}_{j}} \right) \right) \exp\left(-\tau t_{3} \right) \right]. \end{split}$$

Integral:
$$I_4^j(t_1, t_2, t_3) = \int_{t_1}^{t_2} \int_{t_3}^x p_{g_j}(z) \int_0^z p_{g_{(1)}^*g_{(2)}^*}(x, y) dy dz dx.$$
 (9.10)

For Rayleigh fading channel using Eqs. (4.1) and (4.14), the integral in Eq. (9.10) can be evaluated as:

$$\begin{split} I_{4}^{j}(t_{1}, t_{2}, t_{3}) &= \int_{t_{1}}^{t_{2}} \int_{t_{3}}^{x} p_{g_{j}}(z) \int_{0}^{z} \sum_{k,k \neq j}^{K} p_{g_{k}}(x) \sum_{l=1,l \neq j,k}^{K} p_{g_{l}}(y) \prod_{m=1,m \neq j,k,l}^{K} P_{g_{m}}(y) dy dz dx \\ &= \int_{t_{1}}^{t_{2}} \int_{t_{3}}^{x} p_{g_{j}}(z) \sum_{k,k \neq j}^{K} p_{g_{k}}(x) \int_{0}^{z} \sum_{l=1,l \neq j,k}^{K} p_{g_{l}}(y) \prod_{m=1,m \neq j,k,l}^{K} P_{g_{m}}(y) dy dz dx \\ &= \int_{t_{1}}^{t_{2}} \int_{t_{3}}^{x} p_{g_{j}}(z) \sum_{k,k \neq j}^{K} p_{g_{k}}(x) \sum_{\tau \in T_{j,k}^{K}} \operatorname{sign}(\tau) \exp(-\tau z) dz dx \\ &= \int_{t_{1}}^{t_{2}} \sum_{k,k \neq j}^{K} p_{g_{k}}(x) \sum_{\tau \in T_{j,k}^{K}} \operatorname{sign}(\tau) \int_{t_{3}}^{x} p_{g_{j}}(z) \exp(-\tau z) dz dx \\ &= \int_{t_{1}}^{t_{2}} \sum_{k,k \neq j}^{K} p_{g_{k}}(x) \sum_{\tau \in T_{j,k}^{K}} \operatorname{sign}(\tau) \int_{t_{3}}^{x} \exp\left(-\frac{(\tau \bar{g}_{j} + 1)z}{\bar{g}_{j}}\right) dz dx \\ &= \sum_{k=1,k \neq j}^{K} \sum_{\tau \in T_{j,k}^{K}} \frac{\operatorname{sign}(\tau)}{(\bar{g}_{j}\tau + 1)} \left[\left(\exp\left(-\frac{t_{1}}{\bar{g}_{k}}\right) - \exp\left(-\frac{t_{2}}{\bar{g}_{k}}\right) \right) \exp\left(-\frac{(\tau \bar{g}_{j} + 1)t_{3}}{\bar{g}_{j}}\right) \\ &- \frac{g_{j}}{(\tau \bar{g}_{j} \bar{g}_{k} + \bar{g}_{k} + \bar{g}_{j})} \left[\exp\left(-\frac{(\tau \bar{g}_{k} \bar{g}_{j} + \bar{g}_{k} + \bar{g}_{j})t_{1}}{\bar{g}_{k} \bar{g}_{j}}\right) \right]. \end{split}$$
(9.11)