BER-DRIVEN RESOURCE ALLOCATION FOR MULTIMEDIA COMMUNICATION OVER DOWNLINK OFDMA NETWORKS

THAM MAU LUEN

THESIS SUBMITTED IN FULFILLMENT
OF THE REQUIREMENTS
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

FACULTY OF ENGINEERING
UNIVERSITY OF MALAYA
KUALA LUMPUR

2015

UNIVERSITY OF MALAYA

ORIGINAL LITERARY WORK DECLARATION

Name of Candidate:	THAM MAU LUEN (IC/Passport No.:
Registration/Matric No.:	KHA100062
Name of Degree:	Doctor of Philosophy (Ph.D)
Title of Thesis:	BER-DRIVEN RESOURCE ALLOCATION FOR MULTIMEDIA COMMUNICATION OVER DOWNLINK OFDMA NETWORKS
Field of Study:	Wireless Communications

I do solemnly and sincerely declare that:

(1) I am the sole author/writer of this Work;
(2) This work is original;
(3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract front, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work;
(4) I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;
(5) I hereby assign all and every rights in the copyright to this Work to the University of Malaya ("UM") who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;
(6) I am fully aware that if in the course of making this Work I have infringed any copyright whether intentionally or otherwise I may be subject to legal action or any other action as may be determined by UM.

Candidates Signature Date:

Subscribed and-solemnly declared before,

Witness's Signature Date:

Name:

Designation:

ABSTRACT

Recent years have witnessed the growing pervasiveness of multimedia services due to significant advancements of both wireless access and multimedia compression technologies. On one hand, orthogonal frequency-division multiple access (OFDMA) is the key enabler for high-speed multiuser communications. On the other hand, nonscalable and scalable multimedia coding standards offer superior compression gains as well as strong adaptabilities to channel variations. The inherent error-prone nature of channel environments, the wireless resource scarcity along with the variations in the importance among multimedia packets, however, render the quality-of-service (QoS) provisioning for multimedia communication over wireless networks a challenging task. Leveraging multiple target bit-error-rates (BERs) in concert with resource allocation to optimize the received multimedia quality is largely unexplored. Based on the principle that more important packets should be assigned stricter target BERs, this thesis proposes three BER-driven (BRA) resource allocation methods for transmitting pre-encoded multimedia over downlink OFDMA networks, targeting at generic, nonscalable and scalable bitstreams. For the generic case, each bitstream is associated with a different static target BER. Under this constraint, the impacts on three major classes of resource allocation scheme are analyzed from an optimization perspective. For both nonscalable and scalable cases, the target BER of each packet is dynamically adjusted based on scheduling, subcarrier assignment, bit and power allocation, channel quality, and importance level of that packet. The calculation of packet-importance metric takes into account decoder-based error concealment, where both simple and complex techniques are investigated. An additional constraint of strong decoding dependency is considered for the scalable case. Simulation results show that BRA schemes significantly outperform existing BER-unaware techniques in terms of spectral efficiency, power

efficiency and decoded multimedia quality. Further test evaluates the suitability of equal power allocation which is the common assumption in the resource-allocation literature.

ABSTRAK

Tahun kebelakangan ini telah menyaksikan kebiasaan perkhidmatan multimedia yang semakin meningkat disebabkan oleh kemajuan yang ketara dalam teknologi-teknologi akses wayarles dan mampatan multimedia. Dari satu segi, "orthogonal frequency-division multiple access" (OFDMA) adalah kunci utama untuk komunikasi berbilang pengguna yang berkelajuan tinggi. Sebaliknya, tidak berskala dan berskala multimedia pengekodan piawai menawarkan keuntungan mampatan unggul serta kesesuaian kuat untuk variasi saluran. Sifat kesilapan yang sering wujud dalam persekitaran saluran, kekurangan sumber wayarles bersama-sama dengan perubahan dalam kepentingan di kalangan paket multimedia, bagaimanapun, menjadikan "quality-of-service" (QoS) peruntukan untuk komunikasi multimedia melalui rangkaian wayarles satu tugas yang mencabar. Usaha untuk memanfaatkan pelbagai "bit-error-rate" (BER) sasaran bersama-sama dengan peruntukan sumber untuk mengoptimumkan kualiti multimedia yang diterima belum diterokai secara besar-besaran. Berdasarkan prinsip bahawa paket lebih penting perlu diberikan BER sasaran ketat, tesis ini mencadangkan tiga kaedah BER-didorong peruntukan sumber (BRA) untuk menghantar multimedia yang sudah dikodkan melalui rangkaian pautan turun OFDMA, menyasarkan aliran bit generik, tidak berskala dan berskala. Bagi kes generik, setiap aliran bit dikaitkan dengan BER sasaran statik yang berbeza. Di bawah kekangan ini, kesan kepada tiga kelas utama skim peruntukan sumber dianalisis dari perspektif pengoptimuman. Bagi kedua-dua kes tidak berskala dan berskala, BER sasaran setiap paket diselaraska secara dinamik berdasarkan penjadualan, tugasan subpembawa, peruntukan bit dan kuasa, kualiti saluran, dan tahap kepentingan paket itu. Pengiraan paket-kepentingan metrik mempertimbangkan penyembunyian ralat berasaskan penyahkod, di mana kedua-dua teknik-teknik mudah dan kompleks dikaji. Satu kekangan tambahan tentang

pergantungan penyahkodan kuat dipertimbangkan bagi kes berskala. Keputusan simulasi menunjukkan bahawa skim BRA ketara mengatasi teknik BER-tidak-sedar yang sedia ada dari segi kecekapan spektrum, kecekapan kuasa dan kualiti multimedia yang diterima. Ujian seterusnya menilai kesesuaian peruntukan kuasa yang sama yang merupakan anggapan biasa dalam sumber peruntukan kesusasteraan.

ACKNOWLEDGEMENTS

First and foremost, I would like to express my deepest gratitude to my PhD supervisor, **DR. CHOW CHEE ONN**, who has guided me with his patience, motivation, enthusiasm and expertise. He has always been approachable whenever I need to seek for constructive advice, despite his hectic schedule. Apart from the technical knowledge, I have learned so much from his insightful technical writing skill. Moreover, I will never forget his effort to broaden my research perspective by encouraging me to join various extra-curricular activities such as JASSO Student Exchange Support Program.

Next, my gratitude goes to **ASSOC. PROF. DR. KAMARUL ARIFFIN BIN NOORDIN**, **PROF. DR. DANNY HIN-KWOK TSANG** and **PROF. SZU-LIN SU** for the time and effort that they have invested in examining the thesis.

Besides, thanks are due to **DR. NORRIMA BINTI MOKHTAR**, for recommending me to work under the supervision of **PROF. DR. MASAHIRO IWAHASHI** while I was in Japan. It was a pleasant experience working in Iwahashi Laboratory, Nagaoka University of Technology, Japan. I am sincerely grateful to Sensei for teaching me all his knowledge related to image processing. I must also thank all the members of Iwahashi Laboratory for being kind to me, especially to **MS. CHEW YIN PING** from Malaysia, **OGAWA-SAN** and **IKARASHI-SAN** from Japan, and **DR. SUVIT** and **MR. TONSUDTEEN** from Thailand.

My next acknowledgment goes to **PROF. DR. HIROSHI ISHII** for sharing all his research experience. I also thank my colleague, **DR. XU YI HAN** for the stimulating discussions, which are reflected in this thesis results. Besides, I thank **DR. JEEVAN A/L KANESAN** for enlightening me the first glance of mathematical optimization. Of course, it would have been a lonely lab at Level 4, Block R without my lively lab mates,

especially **MS. NANDINI VITEE, MR. NEW WEE KIAT and MR. TAN CHIN YEW**. Besides, to **MS. SARA TAN SU EE**, thank you for your patience on my moody days.

The financial support of **MYBRAIN15 MYPHD** is gratefully acknowledged.

Words are inadequate to express my appreciation to my mentor, Uncle **POH SIN CHEW**, whom I met during one of the most difficult moments in my life. At that time, my self-confidence slumped to the lowest point as I had been facing personal and academic issues. Discerning my anxiety and depression without actually asking, he told me that I could reach out to him anytime. Ever since, he has been emotionally supportive during my good and bad times.

Finally, I owe the most overwhelming debt to my beloved parents, **THAM YAP MEE** and **PANG KING UNG**. Their unconditional love, sustained understanding and endless patience have propelled me to reach the current milestone in my life. I also thank my sister, **SHI YI**, elder brother, **INSP. THAM**, sister-in-law**, INSP. HENG**, and Uncle **TONY** for their invaluable support.

Alan Tham

FEB 2015

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

3G	third generation
4G	fourth generation
AWGN	additive white Gaussian noise
ARQ	automatic repeat-request
ASK	amplitude shift keying
ASNR	average SNR
bpp	bits per pixel
B	bi-directional predictive
BER	bit error rate
BICM	bit-interleaved coded modulation
BRA	BER-driven resource allocation
BRA-SPM	BRA for Sum-Power Minimization
BRA-SRM	BRA for Sum-Rate Maximization
BRA-SUM	BRA for Sum-Utility Maximization
BS	base station
CABAC	context adaptive binary arithmetic coding
CAVLC	context adaptive variable length code
CB	code-block
CDMA	code division multiple access
CNR	channel-to-noise ratio
CP	coding pass
CSI	channel state information
DCT	discrete cosine transform
DWT	discrete wavelet transform
EBCOT	embedded block coding with optimized truncation
EPA	equal power allocation

ESD	entire slice decoding
EXP	exponential
fps	frames per second
FDM	frequency-division multiplexing
FEC	forward error correction
FFT	Fourier fast transform
FPA	frequency-selective power allocation
FSK	frequency shift keying
GOPs	group of pictures
GSM	Global System for Mobile Communications
HH	high-pass filtered in horizontal direction and high-pass filtered in vertical direction
HL	high-pass filtered in horizontal direction and low-pass filtered in vertical direction
HOL	head-of-line
I	intra
IFFT	inverse fast Fourier transform
ISI	intersymbol interference
J2K	JPEG 2000
JA	joint-allocation
JPEG	Joint Photographic Experts Group
JSCC	joint source/channel coding
JSCD	joint source-channel decoding
kbps	kilobits per second
LH	low-pass filtered in horizontal direction and high-pass filtered in vertical direction
LI	less important
LL	low-pass filtered in horizontal and vertical directions
LLR	log-likelihood ratio

LOS	line-of-sight
LTE	Long Term Evolution
LTE-A	LTE-Advanced
Mbps	mega bits per second
M-LWDF	modified largest weighted delay first
M-QAM	M-ary quadrature amplitude modulation
MAC	medium access control
MB	macroblock
MI	more important
MPEG	Motion Picture Experts Group
MR	magnitude refinement
MSB	most significant bitplane
MTU	maximum transmission unit
MSE	mean square error
MU-MIMO	multiuser multiple-input-multiple-output
MV	motion vector
NAL	Network Abstraction Layer
NALU	NAL Units
OFDM	orthogonal frequency-division multiplexing
OFDMA	orthogonal frequency-division multiple access
P	predictive
P/S	parallel to serial
PCRD	post-compression rate-distortion
PER	packet error rate
PF	proportional fairness
PPS	Picture Parameter Sets
PSD	partial slice decoding
PSK	phase shift keying

PSNR	peak signal-to-noise ratio
QCI	QoS class identifier
QCIF	quarter common intermediate format
RTP	real-time transport protocol
S/P	serial to parallel
SD	standard-definition
SISO	single-input-single-output
SNR	signal-to-noise ratio
SP	significance propagation
SPS	Sequence Parameter Sets
SVC	scalable video coding
TDMA	time division multiple access
TR	temporal replacement
UDP	user datagram protocol
UEP	unequal error protection
UMTS	Universal Mobile Telecommunications System
VCEG	Video Coding Experts Group
VCI	video content information
VCL	Video Coding Layer
WBRA	without-BER-driven resource allocation
WiMAX	Worldwide Interoperability for Microwave Access

LIST OF SYMBOLS

$a_1\ a_2$	Arbitrary points
$A_c(t)$	Amplitude of a carrier wave
$A_{k,m}$	Constant corresponding to layer m of user k
A_m	Area of packet m
A_1	Area of overlapped region between the current MB and the reference MB in packet m' of the previous frame
A_2	Area of overlapped region between the current MB and the reference MB in packet m'' of the previous frame
b_{max}	Maximum number of bits corresponding to the highest modulation order
$\overleftarrow{b_{rx}}$	Uncoded received bitstream
$\overrightarrow{b_{tx}}$	Uncoded transmitted bitstream
$\mathcal{B}$	Number of input bits for a chancel encoder
B_n	Bandwidth of subcarrier
B_T	Total available bandwidth
$\mathcal{B}/\mathbb{C}$	Code rate
$\mathrm{BER}_{k,m}$	target BER of queue m of user k
BER_{min}	Minimum target BER
$\overleftarrow{c_{rx}}$	Coded received bitstream
$\overrightarrow{c_{tx}}$	Coded transmitted bitstream
C	total number of code-blocks
$\mathbb{C}$	Number of output coded bits for a channel encoder
$D_{k,f,m}^{L}$	Estimated distortion caused by the lost of packet m after the transmission of $\Pi_k(f, m-1)$ with $\bar{\omega}$
$D_{k,f,m,mb}^{L}$	Estimated distortion caused by the lost of packet m of macroblock mb

$D^R_{k,f,m}$	Estimated distortion caused by the transmission of packet m after the transmission of $\Pi_k(f, m-1)$ with $\bar{\omega}$
$D^T_{k,f,m,mb}$	Estimated distortion caused by the transmission of packet m of macroblock mb
$D_{k,max}$	Distortion without decoding any coding pass for user k
e_k	Number of bits needed to transmit the selected packets from queue k
$E_k[Q]$	Expected quality of user k
f	Frame index
f_c	Frequency of a carrier wave
$F_{k,n}$	Function defined in the problem of video transmission in chapter 4
$F_{k,m,n}$	Function defined in sum-rate maximization problem/ sum-utility maximization problem in chapter 3
$g(\boldsymbol{\mu})$	Dual function
$g'_n(\boldsymbol{\mu})$	Dual function corresponding to subcarrier n
$G_{k,m,n}$	Function defined in sum-power minimization problem in chapter 3
$H_{k,n}$	Channel gain for user k on subcarrier n
i	Iteration index
$J_{k,m}$	Number of remaining unassigned subcarriers needed to transmit packet m of user k
k	User index
K	Total number of users
l	Layer index
L	Number of layers of a JPEG 2000 scalable bitstream
$LER_{k,m}$	Layer error rate for layer m of user k
mb	Macroblock index
M_k	Number of queues/packets for user k
M'_k	Number of queues that has been scheduled

$MB_{k,f,m}$	Number of MBs in packet m of frame f of user k
$MV_{mb,j}^{x}$	Motion vector in x direction
$MV_{mb,j}^{y}$	Motion vector in y direction
n	Subcarrier index
N	Total number of subcarriers
$N_{k,m}$	Number of subcarriers assigned to queue m of user k
N_0	Noise power spectral density
O-level	Number of 2D DWT decomposition
$P_{k,max}$	Maximum allowable transmit power that can be allocated to a layer of user k
P_{tot}	Total available transmission power
$Q_{k,m}$	Contribution of packet m of user k to the reconstructed quality
$r_{k,m,n}$	Number of transmit bits per time slot of subcarrier n for queue m of user k
$\tilde{r}_{k,n}$	Video bits that can be transmitted on subcarrier n for user k
r	Rate allocation matrix
$R_{k,m}$	Length of queue/packet/layer m of user k
s	Power allocation matrix
$s_{k,m,n}$	Power allocated on subcarrier n for queue m of user k
t	Time slot index
t_c	Low pass filtering parameter for rate averaging purpose
$T_{k,m}$	Waiting time of packet m of user k
T_{max}	Playback deadline
T_{ofdm}	Duration of a single OFDM symbol
T_{sym}	Duration of a modulation symbol

U^*_{SPM}	Optimal value of sum-power maximization problem at primal domain
U^*_{SRM}	Optimal value of sum-rate maximization problem at primal domain
V^*_{SPM}	Optimal value of sum-power minimization problem at dual domain
$V^*_{SRM/SUM}$	Optimal value of sum-rate maximization problem/ sum-utility maximization problem at dual domain
w	Coding pass index
W_{inter}	Weight factor of inter MB
W_{intra}	Weight factor of intra MB
$W^{max}_{k,c}$	Number of coding passes for code-block c of user k
x	Width of a video frame
$x_{in}(t)$	In-phase component of a modulation
$x[n]$	Discrete version of a transmitted radio signal
$x_q(t)$	Quadrature component of a modulation
$x_{rx}(t)$	Analog version of a received radio signal
$x_{tx}(t)$	Analog version of a transmitted radio signal
$X[k]$	Frequency-domain signal
y	Height of a video frame
$y[n]$	Discrete version of received radio signal
Z	Arbitrary constant
$\alpha_{k,m}$	Coding loss factor related to target BER of queue m of user k $\mathrm{BER}_{k,m}$
$\beta_{k,m}$	Positive constant corresponding to layer m of user k
Θ	Constraint length of a convolutional encoder
$\emptyset(t)$	Phase of the carrier wave
τ	Duration of a single scheduling slot
σ^2	Noise variance

δ^i	Positive step size at the current iteration i
ϵ	Numerical stopping criterion
$\Pi_k(f, m-1)$	Reordered set of packets in queue k such that $\pi_{k,1}(f)$ is the first packet to be transmitted of current frame f for user k.
$\Delta d_{k,c,w}$	Distortion reduction in terms of normalized MSE of the coding pass w of code-block c of user k
$\Delta F_{k,m}^i$	Subgradient of $g(\boldsymbol{\mu})$ for queue m of user k at the current iteration i
$\Delta r_{k,c,w}$	Rate for coding pass w of code-block c of user k
$\mathcal{L}$	Lagrange function
λ	Lagrange multiplier
λ_0	Water-filling level
$\mathcal{O}$	Big O notation
θ_k	Constant corresponding to user k
$\boldsymbol{\rho}$	subcarrier assignment matrix
Ω	Domain
$\rho_{k,m,n}$	Binary inidicator whether subcarrier n is allocated to queue m of user k
$\psi_{k,m}$	Set of subcarriers assigned to queue m of user k
$\boldsymbol{\tau}$	Set of transmission policy for all users
$\boldsymbol{\tau}_\mathrm{k}$	Vector of transmission policy for all packets for user k
$\tau_{k,m}$	Binary indicator whether packet m of user k is scheduled
Γ	set of available modulation modes
$\Gamma_{k,n}$	modulation constellation size on subcarrier n for user k
$\mu_{k,m}$	Lagrange dual variable for queue m of user k
$\boldsymbol{\mu}$	Lagrange dual variable matrix
$\omega_{k,m}$	Number of successfully decoded MBs for each transmitted packet m of user k

$\bar{\omega}$	Set of overall number of correctly decoded MBs for each transmitted packet from first frame to the previous frame $f-1$
$\boldsymbol{\omega}$	Matrix of overall number of correctly decoded MBs for each transmitted packet
$\gamma_{k,n}$	Channel-to-noise ratio for user k on subcarrier n

CHAPTER 1

Introduction

1.1 Motivation

With around seven billion mobile subscribers in the world nowadays (ITU), there is a fast growth of mobile data traffic owing to an overwhelming demand for mobile services (e.g. multimedia applications, web browsing, file transfers). As shown in figure 1.1, it is envisioned that overall mobile data traffic will rise to 15.9 exabytes per month by 2018, a factor of 8x increment over 2014 (Cisco). Much of this volume of data traffic stem from multimedia content (particularly image and video), which consumes data in the order of at least a few hundred kilobits per second (kbps). The aim to support these bandwidth-intensive applications has spawned an evolution in cellular standards since the introduction of the Global System for Mobile Communications (GSM) in 1991. The GSM successor, Universal Mobile Telecommunications System (UMTS) which is classified as third generation (3G) system, was launched in 2001 in order to achieve even greater data speeds. The fourth generation (4G) technology is the latest standard to arise out of this endeavour.

Unlike its predecessors, 4G systems such as Long Term Evolution (LTE)-Advanced and Worldwide Interoperability for Microwave Access (WiMAX) fully operate in a packet-oriented mode, i.e., all data are segmented into packets which can be transmitted separately. The transmission technology that lies at the heart of 4G systems is orthogonal frequency-division multiplexing (OFDM). The idea of OFDM is to partition a single-user high-rate data stream into a number of parallel lower rate streams, which are then simultaneously transmitted on a number of orthogonal subcarriers. Each subcarrier occupies a narrowband channel, thus increasing the symbol duration. In this sense, the intersymbol interference (ISI) can be significantly alleviated. To further

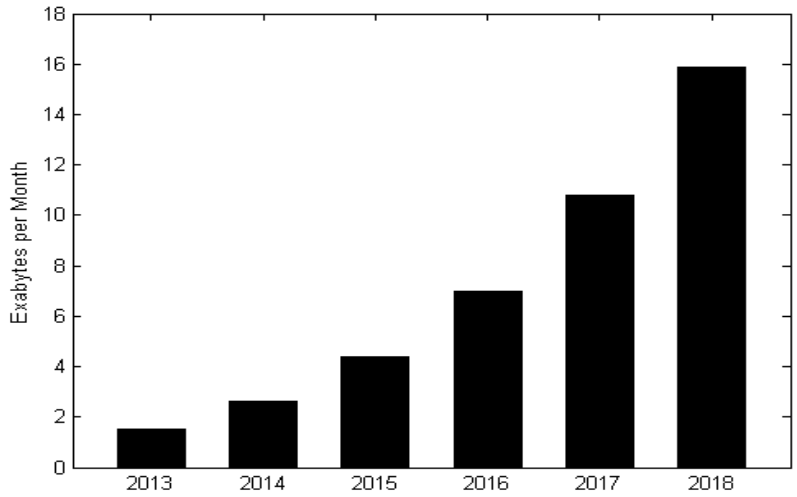


Figure 1.1 Cisco Forecast of Mobile Data Traffic by 2018 (Cisco).

combat ISI, a cyclic prefix is added to the start of each OFDM symbol (van de Beek et al., 1999).

As mobile operators migrate towards serving a large number of subscribers concurrently, orthogonal frequency-division multiple access (OFDMA) goes beyond the OFDM system by scheduling multiple users to share an OFDM symbol, in which different users can be dynamically allocated different sets of subcarriers. Since the channel variations across users are mutually independent, higher spectral efficiency can be achieved by allocating each subcarrier to the user with favourable channel gain on it as compared to static subcarrier assignment. This principle is generally termed as multiuser diversity, performance of which increases with the number of users (Sadr, Anpalagan, & Raahemifar, 2009). More precisely, the performance gain is attributed to the higher degree of flexibility in allocating limited radio resources such as transmission power, subcarriers and bits. Therefore, resource-allocation strategies play a pivotal role in boosting the performance of the OFDMA systems.

There has been a wealth of research work devoted to efficient OFDMA resource allocation schemes. These works often fall into one of three categories, namely: sum-power minimization (Liang & Zhu, 2003; C. Y. Wong, Cheng, Ben Letaief, & Murch,

1999; G. D. Zhang, 2004), sum-rate maximization (Jang & Lee, 2003; Shen, Andrews, & Evans, 2005) and sum-utility maximization (G. C. Song & Y. Li, 2005; Yen, Chang, & Wang, 2010). Transmission power, transmission rate, and fairness among users constitute the common resource constraints. Benefiting from the properly formulated constrained optimization problems in every class above, these solutions intelligently dictate how the scarce radio resources are fully utilized. However, traditional resource allocation approaches treat multimedia bitstream in a data-agnostic fashion (Khalek, Caramanis, & Heath, 2012). That is, the multimedia traffic is simply modelled as generic data with relatively high data rate. In light of this, these schemes regard that throughput maximization will directly translate into improved performance in terms of received multimedia quality, which is not always true (Luo, Ci, Wu, Wu, & Tang, 2010).

1.2 Challenges for Multimedia Communication over Wireless Networks

Unlike generic data, multimedia data demand special treatment not only because of the massive source rate, but also due to their unique attributes (Pande, Ahuja, Sivaraj, Baik, & Mohapatra, 2013). First, different multimedia segments exhibit different rate-distortion characteristics; for example, there is highly a burst source rate from one scene to another scene, depending mainly on the level of spatial-temporal activity. Thus, it is difficult to capture such behaviour within a generic version of multimedia bitstream. Second, multimedia data are very sensitive to jitter and delay. Jitter represents variance of network delay caused by time-varying network conditions. The adoption of a playout buffer can offset jitter, but this may not be feasible in the case of real-time multimedia services such as video conferencing because stricter delay bounds have to be imposed for real-time playback without any interruption. Any packet that does not arrive within its corresponding deadline is considered lost since it is futile for display purpose.

Third, the time sensitiveness of multimedia packets may also render the retransmission of corrupted packets impractical. In light of almost inevitable wireless channel errors, error resilience techniques are important components for alleviating the effect of errors on the decoded multimedia. Encoder-based error resilience techniques (e.g., reversible variable length coding and data partitioning (Moccagatta, Soudagar, Liang, & Chen, 2000)) add redundancy into the compressed multimedia bitstream such that the errors can be detected, localized, or possibly corrected. Decoder-based methods which do not introduce such overhead, attempt to mask the visual artifacts by estimating the missing data from the surrounding decoded data.

Fourth, the encoded multimedia data typically exist in one of two modes: nonscalable and scalable, where this property should be exploited to design efficient resource allocation schemes (Chakareski & Frossard, 2006; Pahalawatta, 2007). A nonscalable coding compresses input multimedia into a single layer with a target bit rate (Sullivan & Wiegand, 2005). To serve a wide range of heterogeneous devices over heterogeneous networks, the encoder must generate multiple versions of the same content over a variety of target bit rates. This, however, may impose significant memory constraints on the encoder side with very large multimedia repositories, considering that there has been a rapidly growing trend of heterogeneity in cellular networks (Q. Y. Ye et al., 2013). A scalable coding is a promising solution in such circumstance (S.-H. Chang, 2011). The idea is to encode the multimedia into multiple layers: a base layer and several enhancement layers. The base layer must be decoded first to render a coarse (but decent) visual quality, whereas every successive enhancement layer refines the quality to a certain degree. This enables senders to optimize the number of transmitting layers that fits well with each device and network capabilities (or requirements), with only one-time encoding.

Such scalability is already supported by preliminary multimedia coding standards such as MPEG-4 Visual ("ISO/IEC 14492-2 (MPEG-4 Visual)," 2004). However, the fact that scalable coding incurs higher complexity and lower coding efficiency over nonscalable coding has hampered their widespread adoption (Schwarz, Marpe, & Wiegand, 2007). Much recent effort has been invested in JPEG 2000 ("ISO/IEC 15444," 2004) and scalable video coding (SVC) in H.264/AVC ("ISO/IEC 14496-10/2005," 2007) in order to bridge the gap in coding efficiency. Despite the evolvement, the seemingly inevitable high complexity of scalable decoder remains an open issue (Hu, Wang, Kwong, Zhao, & Kuo, 2011). On the other hand, the task of decoding a nonscalable (single-layer) bitstream is admittedly much easier, but such too straightforward approach may culminate in memory saturation as discussed above. In brief, there is no single coding method performing best in all situations (Kondrad, Bouazizi, Vadakital, Hannuksela, & Gabbouj, 2009) and it is likely that these two modes will coexist (H. F. Sun, Vetro, & Xin, 2007).

Fifth, irrespective of the multimedia coding mode, different segments of a bitstream entail different levels of importance in terms of reconstructed multimedia quality. For example, in a nonscalable video bitstream, intra (I) data units are more important than predictive (P) data units, which are more critical than bi-directional predictive (B) data units. The granularity level of data unit can range from frame, down through slice to macroblock (MB). In a scalable bitstream, layers are naturally organized based on their importance levels in a hierarchical fashion, where the base layer is more important than the enhancement layers. This property of unequal importance can be exploited in conjunction with unequal error protection (UEP) for efficient multimedia transmission over noisy environments (Maani & Katsaggelos, 2010), which is typically the case in wireless networks. By mapping higher priority (protection) levels to more important

parts, significant multimedia quality gains can be accrued without violating relevant resource constraints (Sabir, Bovik, & Heath, 2010).

1.3 Cross-Layer Design for Multimedia Communication over Wireless Networks

Generally speaking, UEP can be viewed as a form of cross-layer mechanisms, which involve information exchange across functionally separated protocol layers to boost the end-to-end system performance (Jiang, Zhuang, & Shen, 2005). In the context of multimedia communication, parameters at the application, medium access control (MAC), and physical layers are the prevalent ingredients to be jointly designed. One area that has spurred much research interest is optimal real-time encoding, where the source coding at the application layer as well as modulation and coding and transmission power at the physical layer are jointly adapted to the current channel conditions (Eisenberg, Luna, Pappas, Berry, & Katsaggelos, 2002; Hamzaoui, Stankovic, & Xiong, 2005; Huang, Li, Chiang, & Katsaggelos, 2008; M. D. Li, Chen, & Tan, 2011; Luo et al., 2010; Sabir, Heath, & Bovik, 2009; Zhai, Eisenberg, Pappas, Berry, & Katsaggelos, 2005). This thesis, however, focuses on downlink multimedia transmission where the media server is located remotely from a base station (BS) and thus the encoding parameters cannot be adjusted to channel variations. In particular, we assume that the multimedia bitstreams are pre-encoded and packetized before transmitting to the BS. With limited power budget in every BS (e.g., 43 dBm in a 5-MHz LTE ("3GPP TS-36.942 v.8.1.0," 2008)), it is of vital importance to efficiently determines which packet of which user will be transmitted and which resources (power, bits and subcarrier) will be assigned for that packet transmission.

There is a vast literature on packet scheduling and resource allocation for pre-encoded multimedia transmission over wireless networks. The key is to determine the importance level of each multimedia packet so that packets of greater importance will

be assigned higher transmission priority. In (Chou & Miao, 2006; Khalek et al., 2012; van der Schaar & Turaga, 2007), the authors focus on prioritized packet scheduling and retransmission strategy at the MAC layer. Retransmission, however, may not be viable approach given the time sensitiveness of multimedia packets as stated earlier. Additionally, the focus is restricted to single-user multimedia transmission. Likewise, the use of power allocation for single-user transmission is considered in (Atzori, 2003; Sabir et al., 2010; S. S. Tan, Rim, Cosman, & Milstein, 2008).

Works that do consider multiuser scenario can be found in (Chakareski & Frossard, 2006; F. Li & Liu, 2009). However, resource allocation across multiple users is not considered. Such issue is addressed in (Ji, Huang, Chiang, Lafruit, & Catthoor, 2009; Pahalawatta, Berry, Pappas, & Katsaggelos, 2007). Packet scheduling, however, is completely separated from resource allocation where multimedia content from each user queue dictates that user's transmission opportunity. Such approaches, however, are not optimal from a multimedia quality perspective as the transmission of a multimedia packet with greater importance would not always yield better average quality if that packet is not efficient in terms of resource consumption. It is showed in (F. Li, Ren, & Du, 2012) that a joint design of these two aspects offers better quality.

Despite the significant quality gains, all the aforementioned methods are based on the assumption that information bits are all equally important where only one target bit error rate (BER) constraint is considered. Inefficient resource allocation will occur when these schemes are employed to transmit data streams with multiple target BERs (Zhu, 2012). Multimedia bitstreams belong to this kind of bitstreams owing to spatial-temporal variations. In particular, packets with larger contributions to the reconstructed quality should have lower target BERs than the rest. This concept is consistently realized throughout the BER-driven resource allocation (BRA) schemes developed in this thesis.

1.4 Thesis Objectives

The ultimate objective of this thesis is to enhance the performance of pre-encoded multimedia communication over downlink OFDMA networks. This is attained by jointly considering source coding, error resilience at the application layer, packet scheduling and subcarrier assignment at the MAC layer, as well as multiple target BERs, modulation, channel coding, and transmission power at the physical layer. Our springboard is still the generic version of multimedia bitstream as it is fundamentally important to investigate how multiple target BERs affect the overall OFDMA performance. We then extend the concept of multiple target BERs to both realistic transmissions of nonscalable and scalable bitstreams. In this regard, the objectives under scrutiny are:

- To boost the resource utilization for generic multimedia bitstream transmission over OFDMA networks from a mathematical optimization perspective.
- To maximize the average quality of received nonscalable multimedia bitstreams under radio resource constraints.
- To maximize the average quality of received scalable multimedia bitstreams under radio resource constraints.

1.5 Contributions

The major contributions of this thesis can be summarized as follows:

1. BER-driven Resource Allocation (BRA) Schemes

We develop novel approaches that intelligently and efficiently map multiple target BERs to multimedia packets as a function of their visual importance, transmission power, subcarriers, modulation mode and channel conditions in the context of multiuser

multimedia communication. The BRA schemes depart from the traditional schemes in that those schemes assume only one target BER across all types of multimedia bitstream. We first show that the utilization of multiple target BERs gives rise to higher spectral efficiency and power efficiency. We further demonstrate that the enhanced resource utilization translates successfully into better received multimedia quality, for both nonscalable and scalable cases.

2. Proof of Unequal Subcarrier Assignment

Through mathematical derivations in chapter 3, we provide a proof of the common assumption in literature, that important multimedia bitstreams should be assigned subcarriers with good channel conditions in order to have better protection levels. This proposition sparks the development of low-complexity subcarrier assignment strategy in chapter 4.

3. A Novel *Water-Filling* Strategy

Under some channel conditions, it may be better to send fewer packets so that larger portion of resources can be drawn to maintain the overall transmission reliability. As connection quality improves, the transmission system can sustain a larger number of packets with the same degree of reliability, thus resulting in better received multimedia quality. This concept can be realized by the proposed *water-filling* strategy. Unlike the traditional *water-filling* theorem that *water-fills* over frequency, the new technique pours transmission power over all packets as a function of the importance of packet, the length of packet, and the channel conditions. Legitimately, those packets with zero power assignment will not be transmitted.

4. Equal Power Allocation vs. Frequency-Selective Power Allocation

Under the presence of a realistic wireless multimedia transmission system, we aim to cast some light on the suitability of equal power allocation, which is the most prevalently used alleviation in the resource-allocation literature. In fact, most previous works justify the use of equal power allocation from a theoretical perspective and isolate the influence of channel coding. In particular, validations are made by distinguishing the performance between equal power allocation and frequency-selective power allocation in terms of spectral efficiency, where frequency-selective power allocation (FPA) assigns power to subcarriers as a function of their channel conditions. It is however, not clear, how exactly the theoretical performance gap translates into the video performance gap. Our findings are deemed valuable as channel coding is part and parcel of practical systems.

1.6 Thesis Outline

This chapter has analyzed the motivations behind the rapid growth of resource-allocation techniques in OFDMA systems, particularly considering multimedia applications. Additionally, we presented the unique multimedia characteristics that render the packet scheduling and resource allocation a challenging task. We also briefly highlighted the related works, in which cross-layer is the natural option to exploit the property of unequal importance among multimedia data in order to enhance the performance of multimedia communication. We then listed the contributions made in this thesis. The organization of the rest of the thesis is as follows.

Chapter 2 provides some necessary background on multimedia communication over wireless networks. A high-level wireless multimedia communication system is described, starting from source coding, to channel coding, to digital modulation, to wireless channel, to OFDM systems. Along with the elucidation on each component, the

related works are highlighted as well. We also present the research methodology and some assumptions adopted in this thesis.

Chapter 3 presents three BRA schemes for generic multimedia bitstreams with static target BERs, corresponding to three major classes of resource allocation, namely sum-power minimization, sum-rate maximization and sum-utility maximization. The performances of BRA schemes are evaluated in terms of spectral efficiency and power efficiency. The expectably enhanced theoretical performance, however, may not translate successfully into enhanced multimedia quality. The ensuing two chapters are devoted to answering this open question for nonscalable and scalable bitstreams, respectively.

Chapter 4 presents BRA scheme for realistic nonscalable multimedia bitstreams. The target BER of each multimedia packet is no longer static but dynamic. Therefore, it constitutes one of the optimizing parameters in this chapter. The challenge is further magnified when real source coding and practical modulation and coding schemes are taken into account. A three-level framework consisting of packet scheduling, subcarrier allocation and bit and power allocation is devised. Both simple and complex decoder-based error resilience are considered during an accurate estimation of packet importance. Additionally, the complexity issue of the problem is alleviated by using a suboptimal estimation of packet importance. We further test its performance through the comparison with equal power allocation.

Chapter 5 presents BRA scheme for realistic scalable multimedia bitstreams. Together with the shift in the coding mode, the constraint of hierarchical decoding dependencies entails changes in packet scheduling. In this case, an expected quality rather than deterministic quality at the transmitter side must be adopted during the optimized delivery. The reason is that the amount of radio resources allocated to the base layer

directly affects the quality of all the enhancement layers, the amount of resources given to the first enhancement layer directly affect its successive layers, and so forth. Based on the expected quality, a *water-filling* strategy is proposed to optimize the number of transmitted layers and to compute the target BER of the selected layer concurrently, relying on the current channel conditions and the importance of each layer. To reduce the complexity, a suboptimal subcarrier assignment is also proposed.

Chapter 6 contains conclusions and possible future works.

CHAPTER 2

Multimedia Communication over Wireless Networks

In this chapter, we present an overview of a wireless multimedia communication system that is relevant to this thesis. The system serves as the basis for the entire dissertation, i.e., the studies concerning different system objectives associated with different types of multimedia bitstream (generic, nonscalable and scalable) in the ensuing chapters adopt such framework. Emphasis here is placed on the transmission aspect of the system, followed by the system assumptions and methodology to implement the system. Along with the description on each system component, related works are summarized in order to give the readers a big picture of the works in this broad area. Figure 2.1 highlights some major high-level components found in a wireless multimedia communication system. The details of each stage are described as follows.

2.1 Source Coding

A video is composed of a series of images (frames) sampled at regular intervals in time (known as frames per second (fps)). Each image is the representation of a sizeable number of spatial samples called pixels. A monochrome image needs just one component per pixel to represent the brightness of the image whereas a colour image requires at least three components for each pixel such as RGB and YUV. The intensity

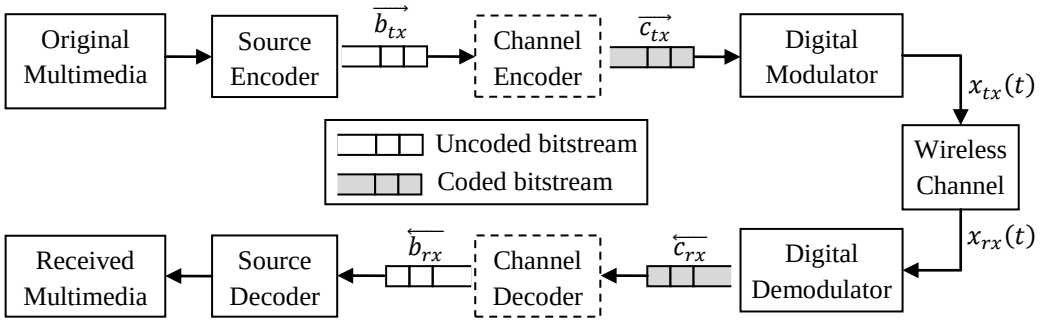


Figure 2.1 Major components in a wireless multimedia communication system.

level that can be represented by each pixel's component depends on the number of bits per pixel (bpp). Consider a standard-definition (SD) colour image of 720x576 8-bit pixels with a sample rate of 25 fps. The transmission rate would require a total of 720x756x25x3x8 ≈ 250 mega bits per second (Mbps). To put this number in perspective, LTE can support a downlink speed of up to 100 Mbps only (Harri & Antti, 2012). Starting from this premise, it is mandatory to compress an original multimedia prior to transmission. This is accomplished by a source encoder that exploits the redundancy in the multimedia and generates a multimedia bitstream $\overrightarrow{b_{tx}}$. As modern communication systems are packet oriented, it is reasonable to assume that $\overrightarrow{b_{tx}}$ consists of a sequence of packets.

2.1.1 Impact of Packet Loss on Received Multimedia Quality

When scheduling and transmitting multimedia packets over wireless networks, it is pervasive that packet loss occurs at the decoder side due to three main reasons. First, some obsolete packets are inevitably dropped at the transmitter owing to scarce radio resources, especially in a multiuser scenario. Second, some transmitted packets may have been distorted and thus discarded. Third, there is a connection disruption when mobile users move from one network to another network. To handle the transmission of multimedia efficiently, it is important to study the impact of packet loss on the received multimedia quality.

Consider a scenario where a user moves from a multicast-capable network to a non-multicast network. Multicast seems to be the best way to deliver multimedia services to a large number of users due to its bandwidth efficiency (Hilt, Sarni, & Lorenz, 2009). While multimedia content delivery could gain performance benefits with multicast, such capability is not consistently available across the entire network infrastructure (Namburi, Sarac, & Almeroth, 2006). For example, although 3G networks support Multimedia

Broadcast Multicast Services (MBMS) for Mobile TV (MTV) services, it is not activated in certain geographical areas or cells (Astrom & Edlund, 2011). This limitation could lead to service disruption when a user moves from a multicast-enabled network to a non-multicast network. Hence, a handover scheme that takes into account unicast and multicast delivery methods is mandatory.

To study the behavior of handover-related packet loss under realistic conditions, we build a real-time streaming testbed as shown in figure 2.2. The testbed is generally divided into three network segments, with network A acting as the server segment, and networks B and C serving as the client segments. Each network is assigned a different static IPv6 address. Two laptops (one for the server and another for the client with two network interfaces) and one desktop computer were used, where each is assigned a

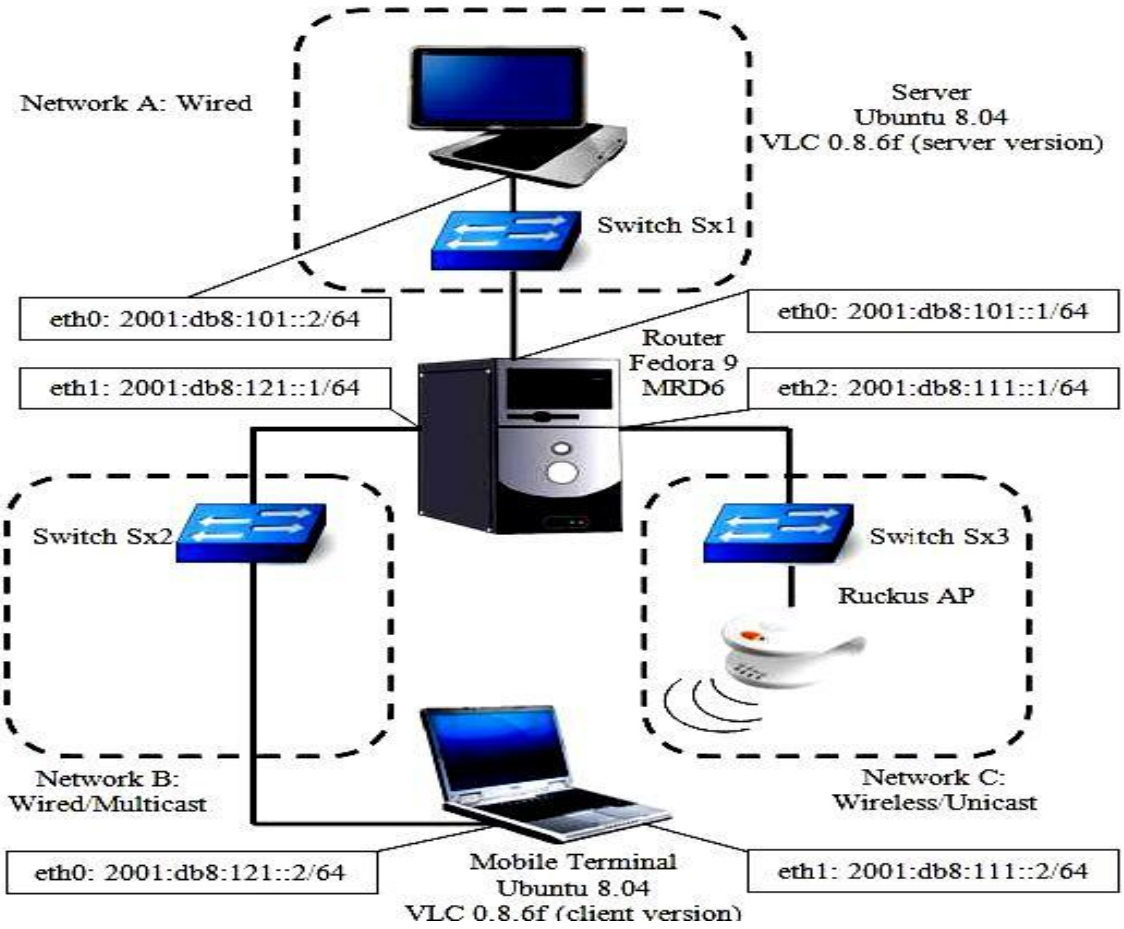


Figure 2.2 Testbed for Seamless Switching between unicast- and multicast- delivered multimedia bitstreams.

static IPv6 address. The Operating Systems (OSs) of the two laptops is Ubuntu 8.04. The desktop computer that acts as the router is installed with MRD6 in Fedora 9 OS with three network interfaces linking all three network segments. MRD6 is an IPv6 routing daemon with multicast forwarding capabilities. The laptop serving as the client is equipped with both Ethernet (802.3) and WLAN (802.11) interfaces, connecting to both the network B and C respectively. The Ethernet is chosen to mimic other cellular networks as we don't have base station equipment. Network C serves as the WiFi access with multicast-capable access point. We adopt the MPEG-2 Transport Stream (TS) for delivering multimedia contents in our experiment. MPEG-2 TS is a popular format for transmission of multimedia streams over networks such as Internet Protocol Television (IPTV).

Three multimedia (A/V) files of different variable bit-rate video and constant bit-rate audio as tabulated in table 2.1 are tested. For all three cases, the target buffer level is set to 300 ms.

Peak signal-to-noise ratio (PSNR) analysis is chosen to study the handover impact on the streamed video as it is often used in literature to study video quality. In short, PSNR is calculated by comparing every pixel in the first frame of the streamed video with the corresponding pixel in the first frame of the pre-encoded video, and similarly to the subsequent frames. If there is no distortion, the PSNR value should be infinity. For

Table 2.1 Properties of Three Test Multimedia Files.

Multimedia	Video fps	Video bitrates (kbps)	Audio fps	Audio bitrates (kbps)
I	15	411	38	128
II	25	770	38	112
III	30	4025	38	160

reason of simplicity, we adopt the same approach as in (A. Chan, Zeng, Mohapatra, Lee, & Banerjee, 2010) to define the highest value of PSNR as 100 dB. The higher the PSNR value, the higher the received frame quality and the higher level of viewing satisfaction experienced by the users.

Figure 2.3 displays the PSNR value of all three videos with and without handover for the selected 100 frames, which shows significant PSNR differences. As expected, for all three videos without handover, the resulting average PSNR value is 100 dB. This means that there are no distortions in any frame of these videos. On the contrary, if there is any frame pair returning a non 100 dB, it turns out to be because of the distortion caused by the handover.

For each streamed video, the first frame pair returning PSNR of non 100 dB has been selected as the starting point of the graph. From the figure, it can be observed that Video II obtains the highest average PSNR among all videos at 97.76 dB with 5 distorted frames. The average PSNR for Video I is 84.02 dB with 26 distorted frames while the average PSNR for Video III declines by 36.3% as compared to Video I with the highest number of distorted frames. Obviously, this indicates that Video III has the lowest QoS

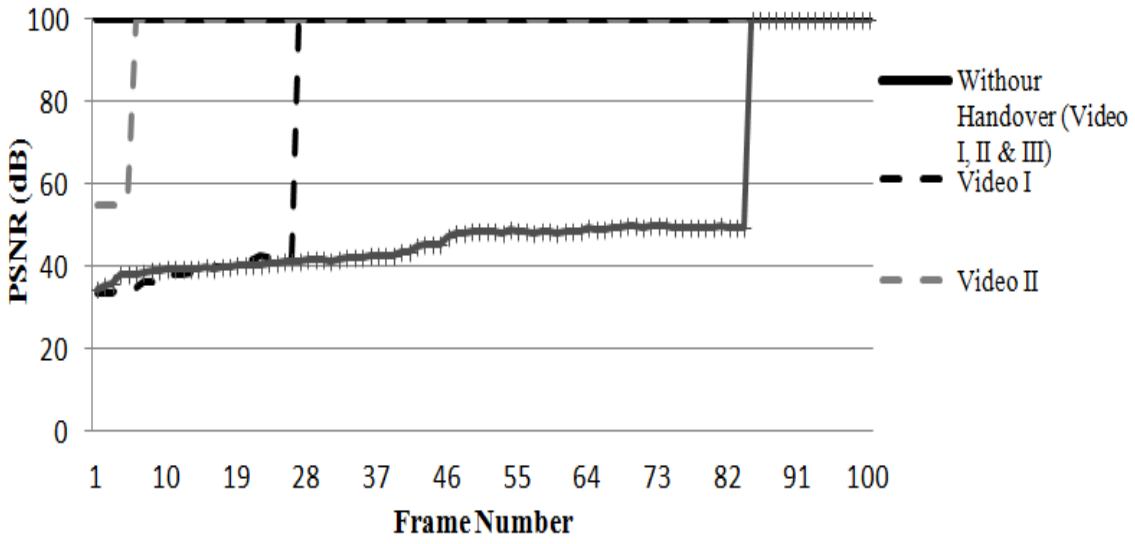


Figure 2.3 PSNR without and with handover.

during the handover. The reason lies in the nature of the predictive video coding technique, more specifically, the Group of Pictures (GOP) structure which results in error propagation (Kiraly, Abeni, & Lo Cigno, 2010).A GOP is a group of successive frames reflecting spatial motion activities in video shots. It always begins with an I-frame which does not require any additional information to reconstruct it and then a P-frame which requires the prior decoding of preceding I- or P-frame in order to be decoded. Therefore, if an error occurs within a GOP (e.g. due to frame lost), the error will propagate till the next reference picture.

2.1.2 Prioritized Packet Scheduling

Inspired by the fact that loss of different packets induces different amount of distortions in the received multimedia, a large pool of literature targeting at prioritized packet scheduling algorithms is available. These schemes can be classified into two major categories: semantic-aware and content-aware. Semantic-aware approaches determine the packet importance at a coarse level. An example is (Tupelly, Zhang, & CHong, 2003) in which the importance levels of video frames are recognized based on frame type (I,P, or B frames) and positions in a group of pictures (GOPs). In a similar concept, the authors in (Klaue, Gross, Karl, & Wolisz, 2003) prioritize intra coded video packets during the transmission.

The second class which works at a finer level is based on distortion estimation. This category can be further split into two small groups: model-based and simulated-based. In the first subgroup, mathematical models are derived to estimate the transmission distortion at a level, ranging from sequence (C. Y. Zhang, Yang, Yu, Yang, & Liu, 2007), through frame (Z. H. He & Xiong, 2006; H. X. Zhang & Rangarajan, 2008) to packet (F. Li & Liu, 2009; Y. F. Zhang, Qin, & He, 2010). This typically involves offline measurement of transmission distortion to predict the model parameters. Also,

the models are often related to the transmission rate such that rate-distortion optimized scheduling strategies can be applied. Examples of such works can be found in (Chakareski & Girod, 2004; Chou & Miao, 2006; Khalek et al., 2012; van der Schaar & Turaga, 2007). These works, however, only consider single-user scenario. The authors in (Chakareski & Frossard, 2006) propose a similar scheduling method in the context of a multiuser scenario.

In the second subgroup, the video encoder endeavours to simulate the decoding behaviours so as to predict the transmission distortion at the pixel level (R. Zhang, Regunathan, & Rose, 2000). In (Pahalawatta et al., 2007), the importance of each video packet is obtained by simulating the decoding behaviour twice in order to acquire the mean square error (MSE) of pixel values between frames decoded with and without the presence of that packet. While this kind of transmission distortion analysis is more accurate, the full decoding of the entire video sequence is always linked with high computational complexity and implementation cost. In general, there is always a tradeoff between the accuracy of importance estimation and complexity. It is, however, not clear, how close the performance of model-based gets to that of simulated-based. Thus, this thesis investigates both model-based and simulated-based content-aware packet scheduling algorithms.

2.1.3 Source-Level Error Resilience

To alleviate the ill-effect of packet loss on the decoded multimedia, it is mandatory to build error resilience into the compressed bitstream. At the source-coding level, error resilient mechanisms can be divided into two groups: encoder-based and decoder-based. Encoder-based error resilient technique usually adds redundancy into the compressed bitstream in order to prevent error propagation. This technique is usually composed of data partitioning, resynchronization marking, and optimal real-time video encoding. In

(van der Schaar, Andreopoulos, & Hu, 2006), the authors partition a scalable bitstream into several priority classes in order to improve the overall received multimedia quality. In (Moccagatta et al., 2000), the authors provide a thorough analysis on the performance of resynchronization marker implemented at the JPEG 2000 and MPEG-4 under various channel conditions. In (R. Zhang et al., 2000), an optimal switching between intra- and inter-coded packet based on channel conditions is proposed to limit the error propagation. Similar concept is advocated in (Cote, Shirani, & Kossentini, 2000; D. P. Wu et al., 2000).

Decoder-based methods that are implemented at the source decoder side, on the other hand, do not introduce such coding overhead. Error concealment, under this category, is part and parcel of several image and video coding standards. It endeavours to mask the visual artifacts caused by lost and erroneous packets. The key principle of error concealment is to exploit correlated information that is available at the decoder so that missing pixels can be estimated. A review of existing approaches to error concealment for the image and video standards can be found in (Chiaraluce, Ciccarelli, Gambi, & Spinsante, 2005; Zargari & Fatemi, 2005). The details of source codec together with its error-resilience methods will be discussed in chapters 4 and 5, when necessary for the scenario.

2.2 Channel Coding

Although source coding itself is a powerful tool to achieve robustness against channel errors, it cannot always overcome the abrupt wireless channel conditions. A channel encoder aids in combating channel impairments by adding redundancy to the bitstream. Distinguished from the multimedia redundancy that we wish to eliminate in the previous section, the redundancy here has an explicit goal of detecting and/or correcting bit errors. This type of channel coding is called forward error correction (FEC). Another type

belongs to automatic repeat-request (ARQ) which retransmits an erroneous packet. This feature is exploited in (Begen & Altunbasak, 2007; Han, Kang, Kim, & In, 2013) for wireless video. However, as discussed earlier, the time sensitiveness of multimedia packets may render this solution powerless. Therefore, we resort to FEC and use the terms "FEC" and "channel coding" interchangeably throughout this thesis.

Broadly speaking, there are two classes of FEC: block codes and convolutional codes. Block codes are referred to as $(\mathbb{C},\mathcal{B})$ codes in which it operates on $\mathcal{B}$-bit non-overlapping message blocks and expands each block into $\mathbb{C}$-bit coded blocks with $\mathbb{C}-\mathcal{B}$ extra bits (parity bits). The coding rate is defined as $\mathcal{B}/\mathbb{C}$; the higher the rate, the lower the channel overhead, the weaker the error-correcting capability. If the block codes are systematic, every coded block of $\mathbb{C}$-bit comprises the $\mathcal{B}$ message bits padded by the $\mathbb{C}-\mathcal{B}$ parity bits. Examples of block codes that have been employed for multimedia transmission over wireless networks are Hamming code (Al-Suhail, Wakamiya, & Fyath, 2006) and Reed-Solomon code (Thomos, Argyropoulos, Boulgouris, & Strintzis, 2005). Unlike block codes in systematic form, convolutional codes send only the parity bits of $\mathbb{C}$-bit length and allow the message blocks to be overlapped. The decision on how many past and new bits involved in each encoding operation is determined by a sliding window with a constraint length Θ, as illustrated in figure 2.4. Such code is explored in (Y. Wu, Kumar, Hu, Zhu, & Matyjas, 2014), for the transmission of video packets. A concatenation of all coded blocks is referred to as coded bit sequence $\overrightarrow{c_{tx}}$.

A channel decoder should interpret the most likely multimedia (message) bits that have been conveyed by the source encoder from among the set of all possible transmitted message bits. As convolutional codes are widely used in practical digital communication systems (K. He & Cauwenberghs, 1999), we will restrict our attention to its decoding process. Consider the received convolutional coded sequence $\overleftarrow{c_{rx}}$ which can exist in one of two forms: digitized or undigitized (more details in section 2.3). The

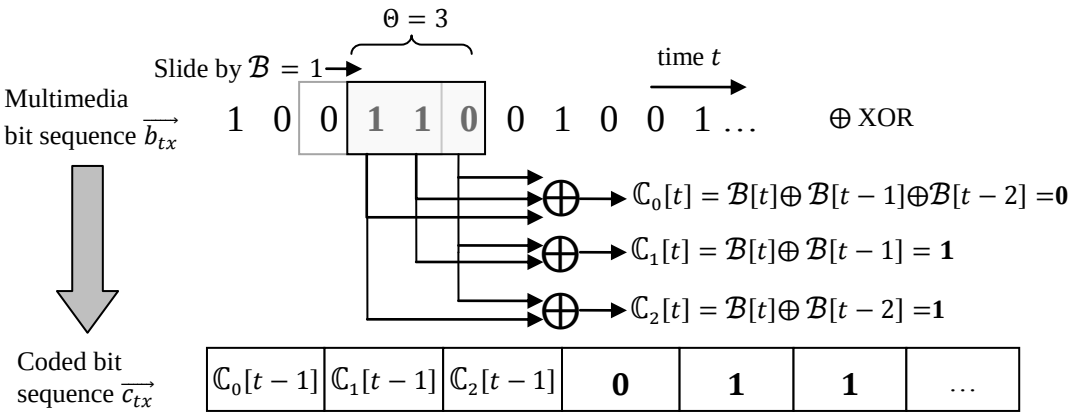


Figure 2.4 An example of a $1/3$ convolutional code with constraint length $\Theta = 3$.

former is simply a hard-decision demodulated bitstream ("0" or "1" in every bit position) while the latter contains soft-decision metric for each bit position to indicate the confidence regarding whether that bit is "0" or "1". For this reason, the channel decoding for digitized and undigitized sequences are termed hard-decision decoding and soft-decision decoding, respectively. Hard-decision decoding expectably yields poorer BER performance than soft-decision decoding as it makes a premature decision about the value of a bit. However, soft-decision decoding is much more complicated to be implemented as it computes soft metric for every coded bit prior to channel decoding. Both decoding modes can be realized with the famous Viterbi algorithm (Viterbi, 1967). This algorithm outputs the received multimedia bitstream $\overleftarrow{b_{rx}}$ with possible unrecoverable bit errors that will be handled by a source decoder.

With proper design, full error correction may be possible given that the amount of bit errors does not exceed the error-correcting capability of a channel code. However, most FEC codes are designed exclusively for good performance in memoryless channels, i.e., bit errors are uniformly distributed (Y. Sun & Xiong, 2006). When error burst occurs, FEC yields poorer practical BER performance than its theoretical counterpart. Time interleaving is an effective tool that spreads the burst error over $\overleftarrow{c_{rx}}$ as shown in figure 2.5. While this can enhance the BER performance of a coded bitstream, its performance

can be worse than that of an uncoded bitstream especially at poor channel conditions (Biglieri, 2005). For this reason, we will consider both uncoded and coded bitstreams for different multimedia transmission cases in the latter chapters. This indicates that channel coding in conjunction with interleaving is optional and explains why the blocks "channel encoder" and "channel decoder" are drawn in dotted line.

Note that this thesis focuses on the link-layer FEC rather than application-layer FEC. In the link layer, redundant bits are added within a multimedia packet to perform intra-packet recovery from bit errors (Zhai et al., 2005). In the application layer, redundant packets are typically generated in addition to source packets to perform cross-packet FEC (Casu, Cabrera, Jaureguizar, & Garcia, 2011). Some cross-packet FEC approaches require modifications to the pre-defined bitstream organization, which may destroy the scalability offered by a multimedia bitstream (Zargari & Fatemi, 2005).

2.2.1 Cross-Layer Design

A step forward from applying FEC alone on source coding is the joint design of these two, which contradicts with the well known Shannon separation theorem (Shannon, 1948). It states that an optimal communications system can be constructed by

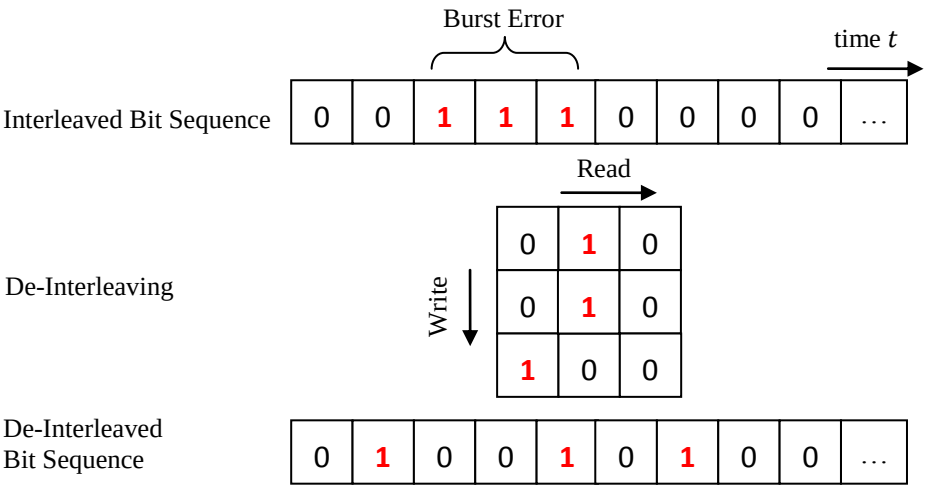


Figure 2.5 An example of time interleaving.

considering source and channel coding separately. The theorem, however, does not hold for realistic systems under the presence of bit errors. As more bits are allocated to source coding, fewer will be left for channel coding, which leads to less bitstream protection. Motivated by this fact, a number of joint source-channel coding methods have been proposed (Hamzaoui et al., 2005; Pejoski & Kafedziski, 2013; Perrine, Chatelllier, Wang, & Olivier, 2008; Shan, 2009; Zhai et al., 2005; X. J. Zhang, Wu, Fowler, Cen, & Dong, 2014). The key idea is to invoke joint source-channel coding, where an optimal bit allocation between source and channel coding are determined at given channel conditions.

In general, techniques that involve information among functionally separated protocol layers can be categorized as cross-layer design. Another popular cross-layer design is joint source coding and power adaptation (Y. S. Chan & Modestino, 2003; Eisenberg et al., 2002; Sabir et al., 2010; Y. H. Tan, Lee, Tham, & Rahardja, 2009), in which the source codec of a packet and the power allocated to that packet are jointly optimized. This thesis, however, focuses on downlink multimedia transmission where a media server is located remotely from the base station and thus the encoding parameters cannot be tuned to channel variations. Despite no interaction between application (source) and the rest layers, our work is still a cross-layer approach as prioritized packet scheduling in section 2.1.1 and dynamic subcarrier assignment (more details in section 2.5.2) at the MAC layer, as well as adaptive modulation (more details in section 2.3.1) and adaptive power allocation (more details in section 2.5) at the physical layer are optimized by considering source codec and source-level error resilience in section 2.1.

2.3 Digital Modulation

As uncoded or coded bitstream is a binary-valued string ("0" or "1"), digital modulation is a natural choice to convert the information bits into an analog form via a sinusoidal carrier wave for transmission purpose. The equation representing a sinusoidal wave with a symbol period of T_{sym} is given by

$$x_{tx}(t) = A_c(t)\cos\big(2\pi f_c t + \emptyset(t)\big) \qquad 0 \le t \le T_{sym} \tag{2.1}$$

where $A_c(t)$, f_c and $\emptyset(t)$ denote the amplitude, frequency and phase of the carrier wave, respectively.

Altering these three characteristics empowers us to map b-bit information data into one of $M = 2^b$ possible symbols (Smithson, 1998). These symbols can be represented in the convenient form of constellation point which consists of in-phase $x_{in}(t)$ and quadrature components $x_q(t)$. This is obtained by rewriting equation (2.1) as

$$x_{tx}(t) = x_{in}(t)\cos(2\pi f_c t) - x_q(t)\sin(2\pi f_c t) = \mathrm{Re}\big[\{x_{in}(t) + jx_q(t)\}e^{j2\pi f_c t}\big] \tag{2.2}$$

where $A_c(t) = \sqrt[2]{x_{in}(t)^2 + x_q(t)^2}$, $x_{in}(t) = A_c(t)\cos\emptyset(t)$ and $x_q(t) = A_c(t)\sin\emptyset(t)$. $x_{in}(t)$,$x_q(t)$ and $\{x_{in}(t) + jx_q(t)\}$ are referred to as in-phase component, quadrature component and complex baseband signal, respectively.

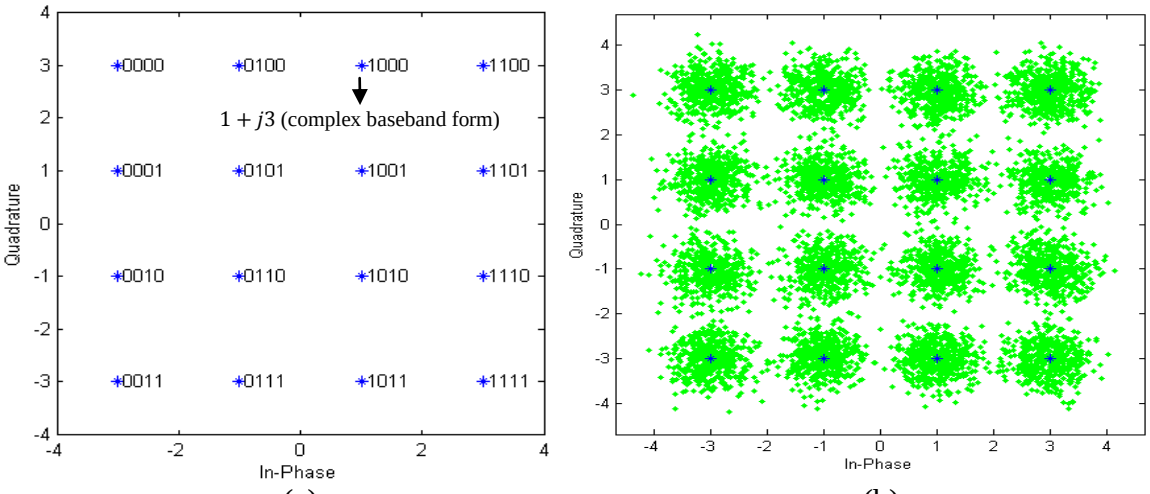
Adjusting only one of three attributes gives rise to one of three modulation forms: amplitude shift keying (ASK), frequency shift keying (FSK) and phase shift keying (PSK). A joint design of these attributes leads to higher spectral efficiency. M-ary quadrature amplitude modulation (M-QAM) which combines ASK and PSK is one of the most widely used digital modulation schemes ("Digital Modulation in Communications Systems –An Introduction," 1997). An example is 16-QAM where 16

possible symbols are created from the combination of four levels of variations in the amplitude and phase as depicted in figure 2.6 (a).

A digital demodulator should recover the signals $x_{in}(t)$ and $x_q(t)$ from the received modulated signal $x_{rx}(t)$. This is attained by multiplying $x_{rx}(t)$ separately with $\cos 2\pi f_c t$ and $\sin 2\pi f_c t$ and applying a low-pass filtering. The operation is mathematically written by

$$\begin{aligned}&\{x_{in}(t)\cos(2\pi f_c t) - x_q(t)\sin(2\pi f_c t)\}\cos(2\pi f_c t)\\ &= x_{in}(t)\cos(2\pi f_c t)\cos(2\pi f_c t) - x_q(t)\sin(2\pi f_c t)\cos(2\pi f_c t)\\ &= 0.5x_{in}(t)[1+\cos(4\pi f_c t)] - 0.5x_q(t)\sin(4\pi f_c t) \qquad (2.3)\end{aligned}$$

With low-pass filtering, only $x_{in}(t)$ remains as the higher frequency components containing $4\pi f_c t$ are eliminated. The $x_q(t)$ extraction is carried out in similar fashion with $\sin(2\pi f_c t)$ acting as the multiplier. As $x_{rx}(t)$ arrives from a noisy environment, the locations of those received symbols (green dots) will not be exactly in the original location of the constellation point as shown in figure 2.4 (b). Based on the received symbol position, either hard-decision demodulation or soft-decision demodulation is



(a) (b)

Figure 2.6 Constellation diagram.
(a) 16-QAM binary symbol mapping (b) 16-QAM received signal.

performed.

Hard-decision demodulation outputs the received signal in bit form (digitized data). On the other hand, soft-decision demodulation computes the probability of a bit having "0" or "1" for each bit position (undigitized data). It is common to employ the log-likelihood ratio (LLR) algorithm (Tosato & Bisaglia, 2002) to compute the soft metric for each bit. Indeed, if the constellation size M is large, there is a high possibility that the demodulator may misinterpret the correct position of the received symbol. This can be alleviated by Gray coding that arranges the signal's constellation diagram such that the bit patterns conveyed by neighbouring constellation points differ by only one bit (Proakis, 1995), as visualized in figure 2.7. Correspondingly, BER performance can be improved.

2.3.1 Adaptive Modulation

To satisfy certain quality-of-service requirements, particularly a target BER, transmission power and modulation constellation size can be jointly adapted according to (instantaneous) channel conditions. Such technique is referred to as adaptive modulation. Assuming square M-QAM is used, the expression for the BER with Gray

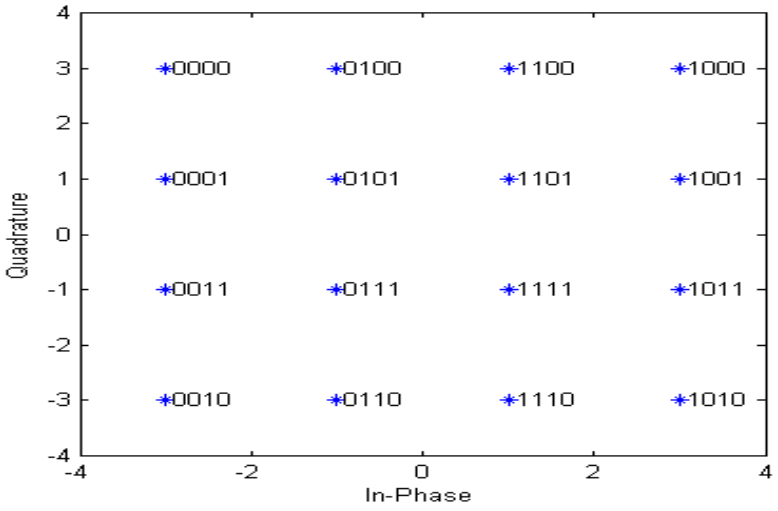


Figure 2.7 16-QAM Gray-coded symbol mapping.

coding as a function of channel-to-noise ratio (CNR) and M is given as

$$\mathrm{BER}_{\mathrm{CNR}} = \frac{2}{b}\left(1 - \frac{1}{\sqrt{M}}\right) \times \mathrm{erfc}\left(\sqrt{1.5\frac{\mathrm{CNR} * s}{M-1}}\right) \tag{2.4}$$

(Chung & Goldsmith, 2001), where b is the number of bits, s is the amount of transmission power, and erfc is the complementary error function. To easily obtain the optimal power and rate adaptation for different modulation schemes, equation (2.4) can be approximated by

$$\mathrm{BER}_{\mathrm{CNR}} \approx 0.2\exp[-1.5\mathrm{CNR}^{*}s/(M-1)] \tag{2.5}$$

(Qiu & Chawla, 1999).

In (Zrae, Hassan, & El-Tarhuni, 2010), an adaptive modulation scheme for image transmission over wireless channel is proposed. It partitions an image into three different quality bitstreams, each with different target BER. For each individual bitstream, constellation size is selected from an offline lookup table according to current channel conditions. In (M. D. Li et al., 2011), another adaptive modulation design is proposed for scalable video transmission over OFDMA networks. The amount of allocated power to match a target BER is determined exactly from equation (2.5). The above two methods, however, assume a predetermined target BER without considering the input multimedia characteristics.

2.4 Wireless Channel Characteristics

Apart from noise which is usually modelled by the additive white Gaussian noise (AWGN), a wireless signal's amplitude (power) attenuates over time and frequency domains. This is called fading which can be categorized into two major classes: large-scale fading and small-scale fading (Goldsmith, 2005), as illustrated in figure 2.8. Large-scale fading arises when a receiver moves long distances. It is caused by the path

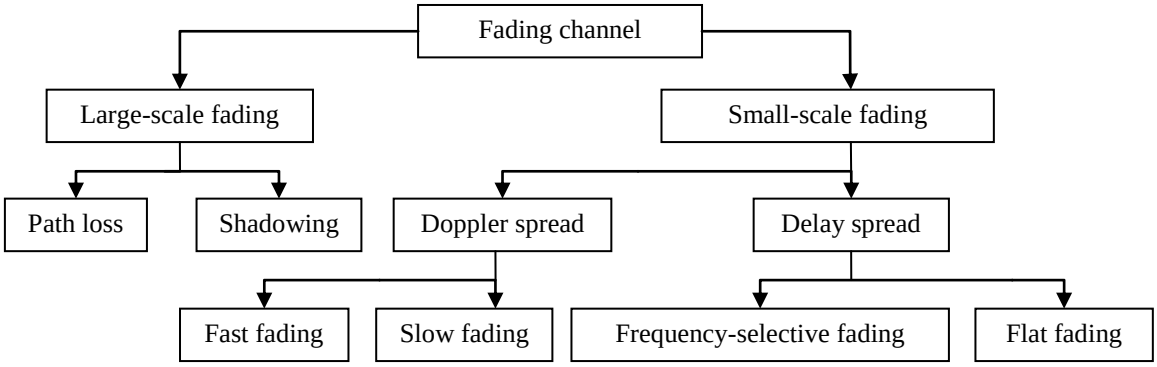


Figure 2.6 Classification of fading channels.

loss of the received signal as a function of distance and shadowing due to non line-of-sight (LOS) propagation environment. Small-scale fading refers to the rapid variation of the received signal amplitude in short time spans as the receiver moves short distances. The amplitude follows Rayleigh and Rician distributions in the non-LOS and LOS environments, respectively. It is due to the presence of multipath signals as a result of the propagation phenomena. When these signals arrive subsequently at the receiver with different phases, constructive or destructive interference occurs. The variation in phase is attributed to two channel parameters: Doppler spread and delay spread.

Doppler spread refers to the range between minimum and maximum Doppler shift encountered over all multipath signals. Doppler shift exerts a deviation in the carrier frequency on each multipath signal, depending on the relative motion between the receiver and that propagation path. Due to Doppler spread, the transmitted signal undergoes time-selective fading either in a fast or slow manner. If the coherence time which is inversely proportional to Doppler spread is smaller than the symbol time T_{sym}, it is subject to fast fading, and slow fading otherwise.

Delay spread is the time spread between the arrival of the first and last multipath signal at the receiver. It defines the coherence bandwidth which determines whether a channel spectral response is flat or frequency-selective. If the coherence bandwidth is larger than

the signal bandwidth, the transmitted signal undergoes flat fading and the channel is called narrowband channel. Otherwise, the transmitted signal undergoes frequency-selective fading and the channel is termed wideband channel. In other words, a wider signal bandwidth implies that the symbol time is much shorter than delay spread. In this case, inter-symbol interference (ISI) occurs as the signals of the current symbol overlap with those of subsequent symbols. ISI becomes very significant in single-carrier transmission systems that supposedly shorten the symbol period for the purpose of high-rate transmission. OFDM systems are well-suited to combat ISI.

2.5 Orthogonal Frequency Division Multiplexing (OFDM)

OFDM is a multicarrier modulation scheme in which multiple carriers (subcarriers) divide the frequency-selective wideband channel into several narrowband subchannels. As long as the bandwidth of each subchannel is smaller than the coherence bandwidth, a data modulation symbol carried onto each individual subcarrier is exposed to flat fading rather than frequency-selective fading. Simply put, the existence of subcarriers alleviates the burden of high data rate that each individual subcarrier must bear, therefore lengthening the symbol interval on each subcarrier. In this regard, the effect of ISI is severely mitigated. Figure 2.9 compares the frequency spectrum of single-carrier versus multicarrier systems.

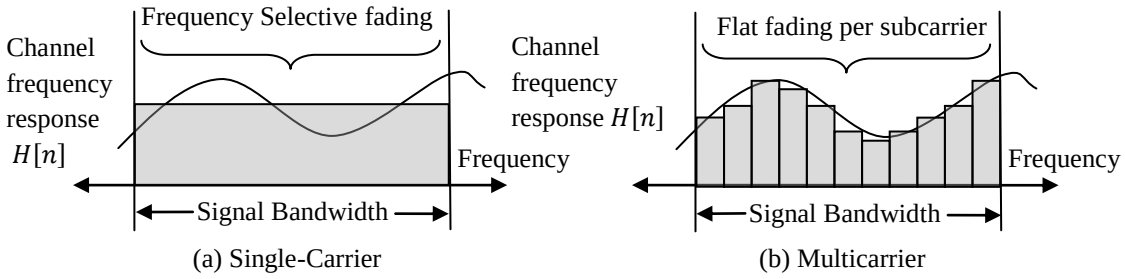


Figure 2.9 Frequency spectrum: single-carrier vs. multicarrier modulation systems.

While the above principle lies at the heart of all multicarrier modulation systems, OFDM distinguishes itself from the conventional frequency-division multiplexing (FDM) systems that divide the channel bandwidth into non-overlapping subchannels. In OFDM systems, the frequency spectrums of adjacent subchannels are allowed to be overlapped by properly choosing each individual subcarrier's frequency such that the subcarrier frequency separation is equal to the reciprocal of the OFDM symbol period T_{ofdm}. This yields to higher bandwidth utilization as compared to those of conventional FDM systems.

Let us consider N discrete-time complex exponential subcarrier signals $\left\{e^{j2\pi f_k nT_s}\right\}_{k=0}^{N-1}$ which correspond to the different subcarriers at $f_k = k/T_{ofdm}$ in the OFDM signal, where $0 \leq t \leq T_{ofdm}$, $t = nT_s = n\,T_{ofdm}/N$, T_s denotes the sampling interval and $n = 0,1, \ldots, N-1$. These signals are orthogonal in that symbol interval if their dot product is zero, that is,

$$\frac{1}{N}\sum_{n=0}^{N-1} e^{j2\pi f_k nT_s} e^{-j2\pi f_i nT_s} = \frac{1}{N}\sum_{n=0}^{N-1} e^{j\frac{2\pi}{N}(k-i)n}$$

$$= \begin{cases} 1, & \forall \text{ integer } k = i \\ 0, & \text{otherwise} \end{cases} \tag{2.6}$$

This results in no interference between the subcarriers and it is possible to extract the subcarrier signal of interest (say f_k) by multiplying the received OFDM signal composing of N data symbols associated with N subcarrier signals by $e^{-j2\pi f_k t}$. In practice, inverse fast Fourier transform (IFFT) and Fourier fast transform (FFT) techniques are implemented to modulate and to demodulate the OFDM symbol, respectively, as illustrated in figure 2.10.

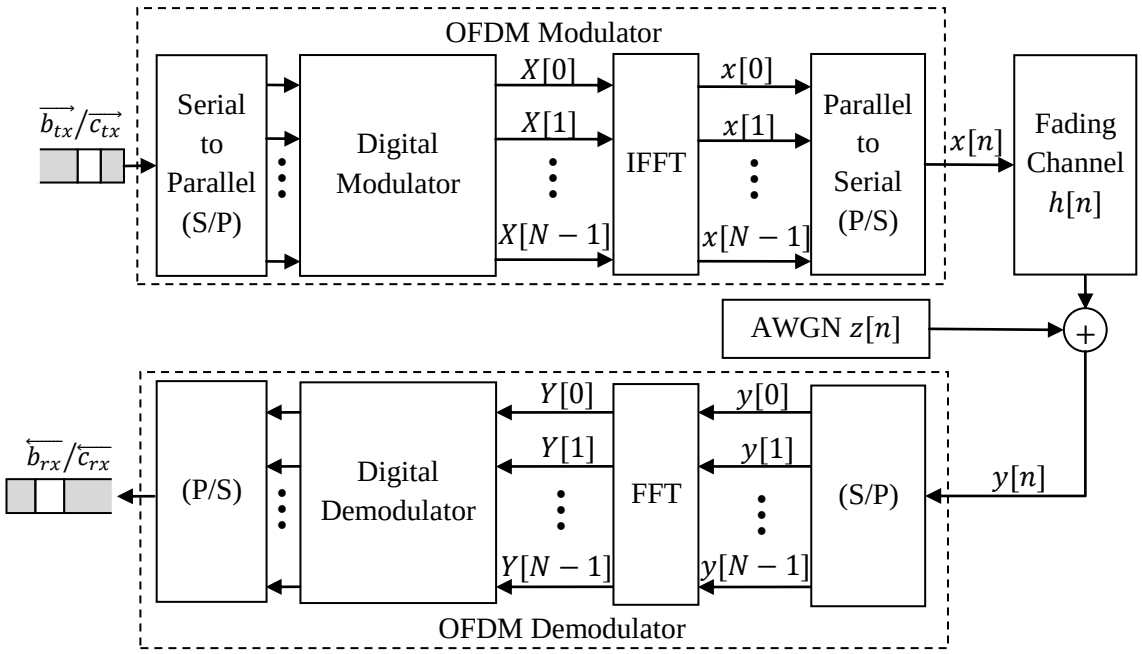


Figure 2.10 Block diagram of an IFFT/FFT based OFDM system.

The input bitstream which can be either uncoded $\overrightarrow{b_{tx}}$ or coded bitstream $\overrightarrow{c_{tx}}$, is first divided into N parallel bitstreams by a serial to parallel (S/P) converter. Each bitstream is then mapped to a constellation point of a standard digital modulation scheme such as M-QAM. When adaptive modulation is employed, different subcarriers can have different constellation sizes, mainly depending on channel conditions of that subcarrier. For example, if the 1st and 2nd subcarriers are associated with 4-QAM and 64-QAM respectively, the first two bits from the bitstream go to the 1st subcarrier while the remaining bits (3rd to 8th) enter the 2nd subcarriers. Note that the means of mapping bits into subcarriers will be totally different if the system is coded OFDM which will be discussed later. The output of the digital modulator is N parallel complex modulation symbols $X[k] = x_{in}(t)_k + jx_q(t)_k$ which will be converted into time-domain symbols $x[n]$ at $t = nT_s$ via an IFFT block. The IFFT operation is mathematically expressed as

$$x[n] = \frac{1}{N}\sum_{k=0}^{N-1} X[k]e^{j\frac{2\pi}{N}kn} \quad \text{for } n = 0,1,\dots,N-1 \tag{2.7}$$

A parallel to serial (P/S) is used to concatenate N parallel $x[n]$ into a single OFDM symbol to be radiated from the transmitter. The transmit signal is then passed through a wireless channel with channel impulse response $h[n]$ (its counterpart is channel frequency response $H[n]$) and AWGN $z[n]$. As a result, the received OFDM signal $y[n]$ at $t = nT_s$ is the combination of the convolution of $h[n]$ and $x[n]$ and $z[n]$, which is given as

$$y[n] = h[n] * x[n] + z[n] = \sum_{m=0}^{\infty} h[m]\, x[n-m] + z[n] \tag{2.8}$$

The OFDM demodulator does the reverse operation to the OFDM modulator. The received OFDM signal is split into N parallel distorted versions of time-domain symbol $y[n]$ which will be converted back to N complex modulation symbols $Y[k]$ through an FFT block. The FFT operation is written as given in (Cho, Kim, Yang, & Kang, 2010):

$$\begin{aligned}
Y[k] &= \sum_{n=0}^{N-1} y[n]\, e^{-j\frac{2\pi}{N}kn} \text{ for } k = 0,1,\dots,N-1 \\
&= \sum_{n=0}^{N-1} \left\{ \sum_{m=0}^{\infty} h[m]\, x[n-m] + z[n] \right\} e^{-j\frac{2\pi}{N}kn} \\
&= \sum_{n=0}^{N-1} \left\{ \sum_{m=0}^{\infty} h[m] \left\{ \frac{1}{N} \sum_{i=0}^{N-1} X[i] e^{j\frac{2\pi}{N}i(n-m)} \right\} \right\} e^{-j\frac{2\pi}{N}kn} + Z[k] \\
&= \frac{1}{N} \sum_{n=0}^{N-1} \left\{ \left\{ \sum_{m=0}^{\infty} h[m]\, e^{-j\frac{2\pi}{N}im} \right\} X[i] \sum_{i=0}^{N-1} e^{j\frac{2\pi}{N}n(i-k)} \right\} + Z[k] \\
&= H[k]X[k] + Z[k]
\end{aligned} \tag{2.9}$$

where $Z[k]$ denotes the noise in frequency domain. The last equality above stems from applying the orthogonality in equation (2.6) where i must be equal to k.

Equation (2.9) implies that the complex modulation symbols $X[k]$ are only affected by its corresponding channel frequency response (channel gain) $H[k]$ and $Z[k]$ at $f_k = k/T_{ofdm}$. This property is very meaningful as it allows to us to adjust the power level of each subcarrier individually so that the channel characteristic as perceived by the user who occupies that subcarrier can be controlled, depending on the objective. If the question is about how to transmit the total data the fastest given a power budget, the *water-filling* theorem (Cover & Thomas, 1991) is the solution which allocates more power to subcarriers with better channel conditions.

At the final stage, all received N parallel complex modulation symbols $Y[k]$ are then baseband-demodulated individually before forming into a serial bitstream $\overleftarrow{b_{rx}}$ / $\overleftarrow{c_{rx}}$ via P/S. Note that when comparing figure 2.10 to figure 2.1, $x_{tx}(t)$ and $x_{rx}(t)$ are actually analog versions of $x[n]$ and $y[n]$, respectively.

2.5.1 Coded OFDM

Channel coding in section 2.2 is also applicable in OFDM systems. While such combination can intuitively boost the BER performance, this is not always the case. The reason is that the BER is relatively high for the subchannels with low channel gains, thereby creating burst errors. Similar to the time-interleaving technique in section 2.2, a frequency interleaving can be applied on the coded bits to spread out the bit errors so that the FEC can perform well (Moon & Cox, 2009). Coded OFDM is a term used for such a system in which OFDM, channel coding and interleaving work closely together. One of the popular coded OFDM systems is based on bit-interleaved coded modulation (BICM) (Caire, Taricco, & Biglieri, 1998) in which the coded bits are interleaved first before mapping onto the subcarriers to achieve frequency diversity. For example, if the 1st and 2nd subcarriers of a single user are associated with 4-QAM and 64-QAM

respectively, the first and the third coded bits go to the 1st subcarrier while the rest is loaded onto the 2nd subcarrier.

2.5.2 Multiple Access Versions of OFDM

The OFDM system discussed so far is applied to a single-user scenario. This implies that OFDM itself is not a multiple access technique, but it can be realized with the combination of existing multiple access schemes such as time division multiple access (TDMA), code division multiple access (CDMA) and frequency division multiple access (FDMA) as depicted in figure 2.11.

An OFDM-TDMA system assigns all subcarriers only to a single user for the interval of several OFDM symbols, in which the number of OFDM symbols can be adjusted in each frame as shown in figure 2.11 (a). In an OFDM-CDMA system, all users are allowed to share both time and subcarriers where a subset of orthogonal codes is assigned to each user and the information symbols are spread in the frequency domain as depicted in figure 2.11 (b). An OFDMA system allows different users to use different subcarriers at the same time as shown in figure 2.11 (c).

Among these methods, OFDMA emerges as the most promising method thanks to its superior BER performance as reported in (Rohling & Gruneid, 1997). This is attributed to the fact that the channel conditions for different users are mutually independent, i.e.,

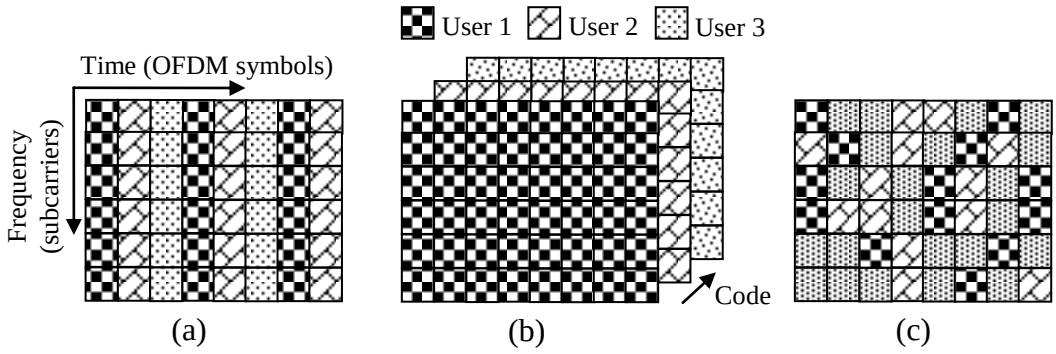


Figure 2.11 Multiple access schemes in OFDM systems.
(a) OFDM-TDMA. (b) OFDM-CDMA. (c) OFDM-FDMA (OFDMA).

the possibility that a subcarrier appears in deep fade for all users is very low. In other words, each subcarrier is likely to be in good channel condition for certain users, especially when the number of users grows (known as multiuser diversity). By dynamically assigning multiple users with good subcarriers, this will translate into better BER performance as compared to those of OFDM-TDMA and OFDM-CDMA systems (G. Li & Liu, 2003).

2.5.3 Perfect Channel State Information (CSI)

As discussed above, one appealing benefit offered by OFDMA is that the multiuser diversity can be leveraged in an optimization process such as *water-filling*. Prerequisite for such optimization is that receiver perfectly estimates and feedbacks its channel state information (CSI) to the transmitter and the CSI is always accessible at the transmitter at the beginning of each transmission interval. The developments done in this thesis are based on the assumption of perfect CSI, which is a common assumption in literature (Sabir et al., 2010).

In practice, it is rarely possible for the wireless transmitter to acquire perfect CSI, due to channel estimation errors and channel mismatch errors. The latter is caused by the delay in channel feedback that results in outdated CSI, considering that wireless channel status is dynamic. To combat feedback delay, several channel prediction methods with reasonable accuracy have been proposed for OFDM systems (Ban, Kim, & Lee, 2008; Jeong & Lee, 2007; Schafhuber & Matz, 2005; I. C. Wong & Evans, 2005). This trend is accompanied by another parallel research line that has investigated the impact of imperfect CSI on the performance of resource allocation schemes (Y. Liu, Ma, & Zhang, 2009; Yao & Giannakis, 2005; S. G. Ye, Blum, & Cimini, 2006). Results show that the gain of adaptive OFDM over non-adaptive OFDM is still significant even in the

presence of imperfect CSI. This implies that the use of adaptive resource allocation in OFDMA systems should be advocated (Sadr et al., 2009)

2.5.4 Packet Scheduling and Resource Allocation for Downlink OFDMA Systems

In a downlink OFDMA system, a base station (BS) transmits data to multiple mobile users. With limited power budget in every BS, it is of vital importance to allocate transmission power to each subcarrier adaptively according to the system objective and constraints. The adaptive power allocation, dynamic subcarrier assignment and adaptive modulation are generally referred to as resource allocation, which is the core research subject in this thesis. Along with the prioritized packet scheduling, we aim to maximize the average received multimedia quality across all users. Figure 2.12 reveals the general form of multimedia communication over downlink OFDMA systems considered in this thesis, in which different multimedia transmission scenarios entail different performance metrics to be evaluated.

2.5.4.1 Generic Case

As generic case does not involve real source codec, power efficiency and spectral efficiency are prevalent performance metrics. Accordingly, there are three major classes of OFDMA resource allocation schemes: sum-power minimization (C. Y. Wong et al., 1999)), sum-rate maximization (Jang & Lee, 2003) and sum-utility maximization (G. C. Song & Y. Li, 2005; Subramanian, Berry, & Agrawal, 2010). One of the main assumptions residing in the above schemes is that all information bits are treated equally. Thus, only one BER constraint is considered. Moreover, these theoretical performances may not reflect the actual received multimedia quality owing to the unique characteristics as discussed in section 1.2.

2.5.4.2 Nonscalable Case

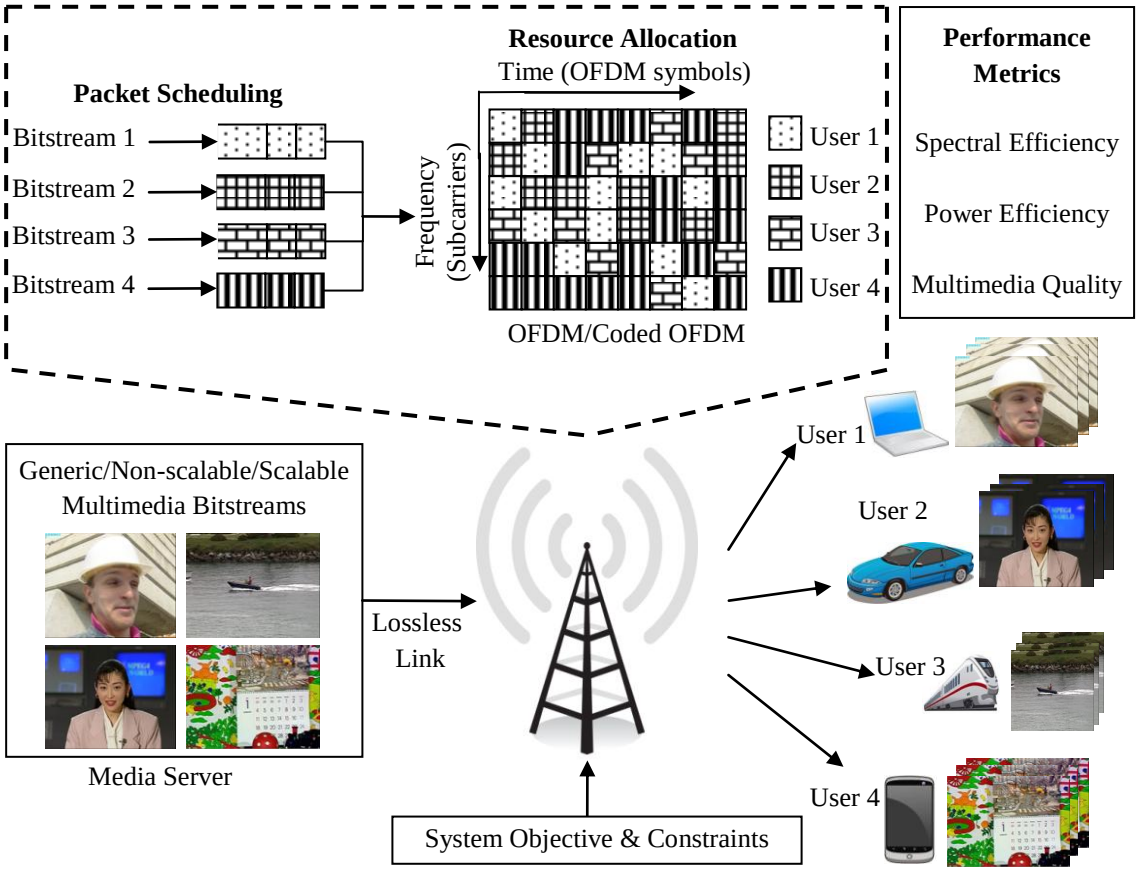


Figure 2.12 The general form of multimedia communication over downlink OFDMA systems considered in this dissertation.

For a nonscalable video bitstream, different video packets possess different levels of importance in terms of reconstructed quality. Inspired by this fact, a large pool of packet scheduling and resource allocation approaches has been proposed to improve the overall video quality. In (Pahalawatta et al., 2007), a scheme that maximizes the weighted sum rate of all video users is proposed. Each user weight is measured by the highest importance level of packet found in that user transmission queue. Similar strategy is advocated in (Ji et al., 2009) for the case of scalable video, except for the additional parameter called packet deadline requirement being included in the weight calculation. The above two schemes, however, share the same drawback, i.e., the users with much higher weights always seize the transmission opportunity. This may lead to resource starvation in the remaining users. Such unfairness issue is addressed in (P. Li, Chang, Feng, & Yang, 2011) by taking into account the scheduled packet size. However,

identical to (Ji et al., 2009; Pahalawatta et al., 2007), the packets are scheduled first by neglecting their resource consumption efficiency. A step forward from allocating resources for a group of scheduled packets is the joint consideration of these two aspects. The authors in (F. Li et al., 2012) demonstrate the promising video performance with such design.

2.5.4.3 Scalable Case

For a scalable bitstream, an inherent packet ordering that restricts the possible scheduling strategies at the scheduler is introduced. That is, every successive enhancement layer contributes a certain amount of temporal/spatial/fidelity refinement only when all preceding layers are reliably received. For example, if the base layer is neither scheduled nor well received, the quality improvements contributed by the successive layers are futile. Extensive research has focused on the resource allocation for scalable-bitstream transmission. In (Park, Lee, Lee, & Bovik, 2009), a cross-layer approach that jointly tunes SVC source and wireless channel parameters is proposed. The authors in (Khalek et al., 2012) proposes another cross-layer design for scalable video over wireless networks. An additional parameter called playback buffer state is taken into account. The above two methods, however, consider only a single-user scenario. While the authors in (Yang & Wang, 2013) do address the relevant problem in the context of multiuser scenario, equal power allocation is adopted

2.6 Methodology and Simulation Setup

The performance evaluation is carried out by means of simulation. To obtain simulation modeling as accurate as possible, we employ several reputable simulation tools in the simulation framework. Basically, the framework comprises two main simulators that are associated with upper (application) layer and lower (MAC and physical) layer, respectively. Note that we do not model the transport and network layers as the work

here concentrates on the last hop of a wireless network where this is likely to be the main bottleneck of the whole multimedia transmission system.

For the upper layer, we adopt real source codec model rather than mathematical model. To this aim, open-source multimedia implementation tools such as JasPer ("JasPer software reference manual," 2002) and JVT ("H.264 JM 18.4 Reference Software,") are used. The lower layer is responsible for channel model, scheduler model, modulation and coding modes and power allocation. The proposed packet scheduling and resource allocation mechanisms reside in the scheduler model. Together with the models associated with the physical layer, they are both implemented by the scientific software Matlab. This is a reasonable choice as we deal with bit-domain operation (Wehrle, Günes, & Gross, 2010). However, software incompatibility issue might arise due to plurality of programming languages and thus an interface should be present. As the chosen upper-level simulators operate on C language, MEX files act as an interface between the upper and lower layers. In particular, the MEX files created in the Matlab assist us in calling C functions in the upper layer.

The general flow of the simulation framework works as follows. First, a number of standard test raw images and videos are collected. Second, depending on the encoding parameters, the upper-level simulator encodes each input multimedia file into a compressed bitstream that consists of a sequence of packets. Third, the packets are loaded into the Matlab. For simplicity, we assume that the application-layer parameters such as the importance and the length of each packet can be accessed at the Matlab so that the proposed schemes take charge. Fourth, the scheduled packets are modulated and sent over the wireless channel. Fifth, the modulated symbols interact with the channel model that accounts for noise and fading. Sixth, the symbols are demodulated to form a received bitstream that may have been distorted. Finally, the packets are passed to the upper-level simulator for decoding and evaluation purposes.

2.7 Multimedia Communications over Legacy Systems (3G)

While multimedia communications is a topic worthy of investigation for both industry and academia communities in 4G cellular systems (Piro, Grieco, Boggia, Fortuna, & Camarda, 2011), extensive research activities regarding multimedia communications have already been carried out for 3G systems. Examples of such works can be found in (P. R. Chang, 1997; Connie, Nasiopoulos, Leung, & Fallah, 2008; Huang et al., 2008; Kambhatla, Kumar, & Lima, 2008; Kim & Kim, 2003; Y. P. Liu, Zhang, Xu, Zhang, & Liu, 2009; Lu, Wang, Erkip, & Goodman, 2007; Pahalawatta et al., 2007; Ukhanova, Belyaev, Wang, & Forchhammer, 2012; Xu & Zhou, 2004). However, these prior works concerning 3G systems cannot be directly applied to the transmission scenario of 4G networks due to completely different multiple access techniques. 4G systems such as LTE rely on OFDMA technology while 3G systems are based on CDMA scheme. Compared to CDMA, OFDMA offers higher degrees of freedom in resource allocation to explore time, frequency and multiuser diversities. At the same time, it makes the optimization task more challenging (Huang, Subramanian, Agrawal, & Berry, 2009).

2.8 Summary

This chapter has provided an overview of multimedia communication over wireless networks with emphasis on the transmission part. Along with the technical description on each major component, we summarized some related works in the literature to ease the understanding of the overall trend in this area. We also listed some assumptions in the system setup and highlighted the new hurdles associated with resource-allocation strategy in 4G systems as compared to those of 3G systems. Armed with such knowledge, we are ready to deal with OFDMA resource allocation.

CHAPTER 3

BER-Driven Resource Allocation (BRA) for Generic Multimedia Bitstreams

In this chapter, we revisit all three major classes of resource-allocation problems from the perspective of optimization, with the consideration of multiple target BERs. Each target BER is associated with one type of generic multimedia bitstream. The optimization problems are formulated with respect to each system objective and constraints. This chapter ends with a simulation study detailing the performances of the proposed schemes in terms of spectral efficiency and power efficiency.

3.1 Introduction

There is a considerable body of research work devoted to designing efficient OFDMA resource-allocation schemes for generic bitstreams. These schemes often fall into one of three categories, namely: sum-power minimization, sum-rate maximization and sum-utility maximization (Sadr et al., 2009). In the first category, the objective is to minimize the total transmission power subject to the minimum data rate requirement for each user, pioneered by (C. Y. Wong et al., 1999). Other examples include (Y. X. Wang, Chen, & Wei, 2005; G. D. Zhang, 2004). In the second category, the goal is to maximize the total data rate (or equivalently, spectral efficiency) of the OFDMA system under the total transmit power constraint. It is proved in (Jang & Lee, 2003) that the highest spectral efficiency is obtained when each subcarrier is allocated to the user with the best channel gain on that subcarrier and transmit power is distributed among the subcarriers using the *water-filling* theorem. However, a key concern arising from this strategy is the unfairness issue, in which the users suffering deep fading for a long period may starve for radio resources. To alleviate this issue, data rate proportionality among the users which acts as a constraint is introduced in the maximization problem (Rhee & Cioffi, 2000; Shen et al., 2005).

In the last category, the utility theory is applied to formulate the resource-allocation problem in OFDMA systems (G. C. Song & Y. Li, 2005). (Subramanian, Berry, & Agrawal, 2010) proves that the maximum sum utility over long term throughout region can be attained by a gradient-based scheduler which maximizes the projection of the achievable data rate vector onto the gradient of a utility function. In other words, the gradient-based scheduler selects the maximum weighted sum rate, where the weights are defined by the gradient of the utility. Depending on the objective of the system, different types of utility functions can be mapped with different performance metrics such as average throughput (G. C. Song & Y. G. Li, 2005), queue length and delay (Song, Li, & Cimini, 2009). Interestingly, if the utility function is chosen to be concave such as logarithmic function, it offers another form of guaranteeing fairness among the users since the slope of the utility function for each user diminishes with an increase in its related performance variables. These types of gradient-based schedulers are said to be proportional fair (H. Kim & Han, 2005).

One of the main assumptions residing in the above schemes is that all information bits are treated equally. Thus, only one BER constraint is considered where the impact of the multiple target BERs is neglected. Indeed, in current-generation wireless networks such as LTE, different target BERs are rendered to different services (e.g. real time gaming and video streaming) (Alasti, Neekzad, Hui, & Vannithamby, 2010). These services conforms to the LTE standardized quality-of-service (QoS) class identifier (QCI) characteristics ("3GPP TS 23.203 v. 8.8.0," 2009) as listed in table 3.1. Note that while only packet error rate (PER) is available in the list, BER can actually be derived from the following equation

$$\mathrm{PER} = 1\text{-}(1\text{-}\mathrm{BER})^{P} \tag{3.1}$$

where *P* is the packet length in bits. Hence, based on QCI and the size of the transmit packet which is bounded by the maximum transmission unit (MTU), it is left open to service providers to define target BER for each traffic class. Obviously, when applying the conventional schemes to transmit heterogeneous traffic flows with distinct BER requirements, inefficient allocation will occur (Zhu, 2012). This is because the selected BER constraint has to be the strictest one. Hereafter we refer the proposed and the conventional schemes to as BER-driven resource allocation (BRA) and without-BER-driven resource allocation (WBRA) schemes, respectively.

The rest of this chapter is organized as follows. Section 3.2 describes our system model. In Section 3.3, we formulate the BER-driven resource allocation (BRA) problems, and present optimal solutions. Section 3.4 provides simulation details and results. We summarize the chapter in section 3.5. A list of the key mathematical notations used in this chapter is provided in table 3.2.

Table 3.1 LTE standardized QoS Class Identifier (QCI) attributes ("3GPP TS 23.203 v. 8.8.0," 2009).

QoS Class Identifier	Packet error rate	Example services
1	10^{-2}	Conversational voice
2	10^{-3}	Live streaming
3	10^{-3}	Real-time gaming
4	10^{-6}	Buffered streaming
5	10^{-6}	IMS signaling
6	10^{-6}	Buffered streaming
7	10^{-3}	Interactive gaming
8	10^{-6}	TCP-based (e.g., ftp, p2p, email)
9		

Table 3.2 List of Key Notations

Notation	Description
PER	Packet Error Rate
P	Packet length in bits
B_T	Total bandwidth
N	Number of subcarriers
K	Number of users
M_k	Number of queues per user
B_n	Bandwidth of subcarrier
$\mathrm{BER}_{k,m}$	Target BER of queue m of user k
N_0	Power spectral density of white noise
$H_{k,n}$	Channel gain on subcarrier n for user k
$\gamma_{k,n}$	Channel-to-noise ratio on subcarrier n for user k
τ	Length of single scheduling slot
$s_{k,m,n}$	Power allocated on subcarrier n for queue m of user k
$r_{k,m,n}$	Number of transmit bits of subcarrier n for queue m of user k
$\alpha_{k,m}$	Coding loss factor related to $\mathrm{BER}_{k,m}$ by adopting MQAM modulation
$\rho_{k,m,n}$	Indicator for subcarrier assignment
$\psi_{k,m}$	Set of subcarriers assigned to queue m of user k
$N_{k,m}$	Number of subcarriers assigned to queue m of user k
$R_{k,m}$	Minimum rate requirements on queue m of user k
P_{tot}	Total transmission power
$w_{k,m}$	Weight factor
$\overline{R_{k,m}}$	Estimate of average throughput for queue m of user k
$g(\boldsymbol{\mu})$	Dual function
Ω	Domain satisfying constraints (3.2) and (3.6)
$\mu_{k,m}$	Lagrange multiplier associated with queue m of user k
δ^i	Positive step size for the update of subgradient method
λ	Lagrange multiplier associated with the power constraint
t_c	Low pass filtering parameter for rate averaging purpose

3.2 System Overview and Problem Formulations

3.2.1 System Model

The block diagram of the downlink OFDMA system under consideration is shown in figure 3.1. In this system with total bandwidth B_T, there are N subcarriers to be shared among K users with M_k queues each. For each user, multiple queues are adopted to store different types of packets associated with different types of multimedia applications. Each multimedia bitstream m of user k is assigned with a different target $\text{BER}_{k,m}$. Each subcarrier occupies a bandwidth of $B_n = B_T/N$ which is assumed to be smaller than the coherence bandwidth in the frequency-selective wireless channel so that each subcarrier undergoes flat fading. The OFDM signaling is time-slotted. The length of each time slot and the length of a single OFDM symbol are denoted by τ and T_{ofdm} respectively. At the receiver, AWGN of power spectral density N_0 with mean zero and variance $\sigma^2 = N_0 B_n$ is present. For user k on subcarrier n, the channel gain and the corresponding channel-to-noise ratio (CNR) are denoted by $H_{k,n}$ and $\gamma_{k,n} = |H_{k,n}|^2/\sigma^2$, respectively. We assume the OFDM system has perfect CSI as stated in

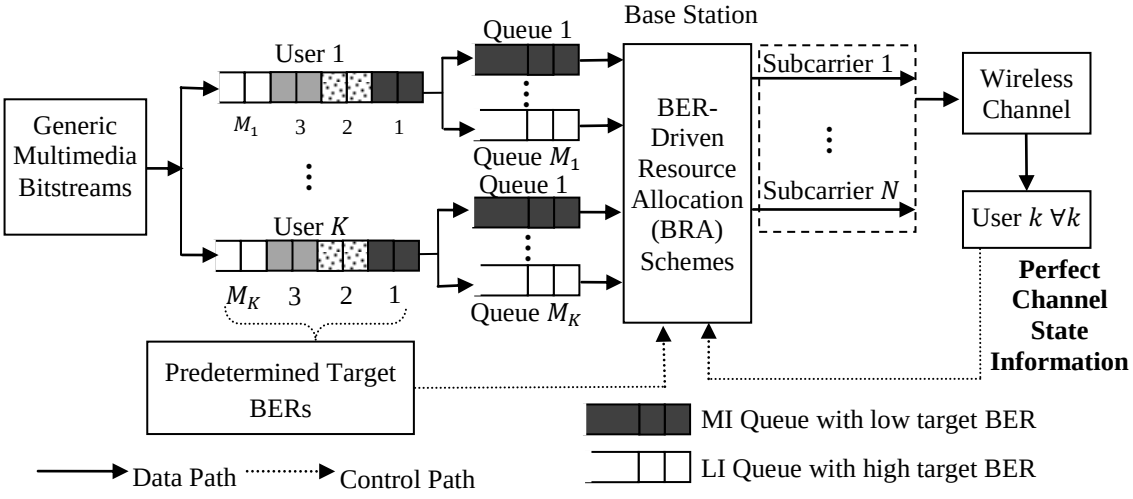


Figure 3.1 System block diagram for downlink OFDMA transmission of generic multimedia bitstreams.

section 2.5.3.

Based on predetermined target BERs and CSI, BER-driven resource allocation (BRA) schemes allocate subcarriers and transmit power to multiple bitstreams once every time slot t. It is assumed $\gamma_{k,n}$ remains static within each slot τ due to slow fading. Denote $s_{k,m,n}$ as the power allocated on subcarrier n for queue m of user k which needs to satisfy

$$s_{k,m,n}[t] \geq 0 \quad \forall k, \forall m, \forall n. \tag{3.2}$$

Then the corresponding number of transmit bits per time slot of subcarrier n for queue m of user k, $r_{k,m,n}$ can be expressed as

$$r_{k,m,n}[t] = \log_2(1 + \alpha_{k,m}\gamma_{k,n}[t]s_{k,m,n}[t])(\tau/T_{ofdm}) \quad \forall k, \forall n. \tag{3.3}$$

where $\alpha_{k,m}$ is a coding loss factor related to target BER of queue m of user k $\mathrm{BER}_{k,m}$ by adopting MQAM modulation (Qiu & Chawla, 1999)

$$\alpha_{k,m} = \frac{1.5}{-\ln(5\mathrm{BER}_{k,m})} \forall k, \forall m. \tag{3.4}$$

Generally, we let

$$\mathrm{BER}_{k,1} \leq \mathrm{BER}_{k,\dots} \leq \mathrm{BER}_{k,M_k} \ \forall k \tag{3.5}$$

where queue m^* is more important (MI) queue while queues $m > m^*$ are less important (LI) queues. From equations (3.3) and (3.4), it is clear that given a fixed amount of transmit power, the number of transmit bits using higher target BER is higher than the one with lower target BER. Let us denote $\rho_{k,m,n}$ as the indicator for subcarrier assignment, where $\rho_{k,m,n} = 1$ implies that the subcarrier n is allocated to queue m of user k; otherwise, $\rho_{k,m,n} = 0$. Since each subcarrier cannot be shared by multiple queues, the binary variable $\rho_{k,m,n}$ should satisfy

$$\rho_{k,m,n}[t] \in \{0,1\}, \sum_{k=1}^{K}\sum_{m=1}^{M} \rho_{k,m,n}[t] = 1 \quad \forall n. \tag{3.6}$$

From here onwards, we omit the time index t for simplicity since it remains unchanged throughout this section. Let $\psi_{k,m}$ and $N_{k,m}$ be the set of subcarriers and the number of subcarriers assigned to queue m of user k, respectively. Then, only subcarrier n with $\rho_{k,m,n} = 1$ is included in $\psi_{k,m}$. For ease of notation, we let overall power allocation as a matrix $\mathbf{s}$ with $[\mathbf{s}]_{k,m,n} = s_{k,m,n}$, overall rate allocation as a matrix $\mathbf{r}$ with $[\mathbf{r}]_{k,m,n} = r_{k,m,n}$, and subcarrier assignment as a matrix $\boldsymbol{\rho}$ with $[\boldsymbol{\rho}]_{k,m,n} = \rho_{k,m,n}$. Armed with above notation, we are ready to formulate the BRA problems.

3.2.2 Problem Formulations

3.2.2.1 BRA for Sum-Power Minimization (BRA-SPM)

The objective is to minimize the total transmit power subject to minimum rate requirements on each queue of all users. The rate constraints can be expressed as

$$\sum_{n=1}^{N} \rho_{k,m,n} r_{k,m,n} \geq R_{k,m} \quad \forall k, \forall m. \tag{3.7}$$

Then the problem can be formulated as

$$U_{SPM}^{*} = \min_{\{\boldsymbol{\rho},\mathbf{s}\}} \sum_{k=1}^{K}\sum_{m=1}^{M_k}\sum_{n=1}^{N} \rho_{k,m,n} s_{k,m,n}$$

$$\text{subject to (s.t.) (3.2), (3.6), (3.7).} \tag{3.8}$$

3.2.2.2 BRA for Sum-Rate Maximization (BRA-SRM)

Here the objective is to maximize the total instantaneous transmission rate subject to a total transmission power P_{tot}.

$$\sum_{k=1}^{K}\sum_{m=1}^{M_k}\sum_{n=1}^{N}\rho_{k,m,n}s_{k,m,n} \le P_{tot} \tag{3.9}$$

Then the problem can be formulated as

$$U_{SRM}^{*} = \max_{\{\boldsymbol{\rho},\mathbf{s}\}}\sum_{k=1}^{K}\sum_{m=1}^{M_k}\sum_{n=1}^{N} w_{k,m}\rho_{k,m,n}r_{k,m,n}$$

$$\text{s.t. (3.2), (3.6), (3.9).} \tag{3.10}$$

where $w_{k,m}$ is a weigh factor that will be set to one. The reason for introducing this parameter will emerge in the following case.

3.2.2.3 BRA for Sum-Utility Maximization (BRA-SUM)

Here the objective is to maximize the total system utility subject to a total transmission power P_{tot}. In this category, we consider proportional fairness (PF) problem in which the total system utility is the sum of logarithmic average user rates. This problem is of particular interest as maximizing solely the transmission rate might lead to unfairness issue. PF scheme ameliorates this issue by maximizing the long-term throughput instead of the instantaneous one. This is achieved by setting the weight factor as

$$w_{k,m} = 1/\overline{R_{k,m}}[t] \quad \forall k, \forall m, \tag{3.11}$$

rather than one as in the BRA-SRM case, where $\overline{R_{k,m}}[t]$ is the estimate of average throughput of queue m of user k at the current time slot t. In other words, BRA-SUM is the same as BRA-SRM except that $w_{k,m}$ is updated at every time slot. For this reason, the optimal solutions for BRA-SRM and BRA-SUM will be discussed in the same subsection later.

Although we focus on the PF problem here, the proposed method can be directly extended to other resource-allocation problems by modifying the setting of $w_{k,m}$ (e.g.,

modified largest weighted delay first (M-LWDF) (Andrews et al., 1998) and exponential (EXP) (Shakkottai & Stolyar, 2002) where the value of $w_{k,m}$ can be found in ((Song et al., 2009), equation (9)). Hereafter the term "SUM" and "PF" will be used interchangeably.

3.3 Optimal Solutions

All three above optimization problems belong to nonconvex optimization problem due to the discrete set of subcarrier assignment. Hence, it is prohibitively complex to obtain the optimal solution $\{\boldsymbol{\rho}^*, \mathbf{s}^*\}$ since there are $(\sum_{k=1}^{K} M_k)^N$ possible subcarrier assignment for an OFDMA system with large K, M_k and N values. Fortunately, this kind of problem can be solved efficiently by using the Lagrange dual decomposition method as long as the number of subcarriers becomes sufficiently large (Huang et al., 2009; Seong, Mohseni, & Cioffi, 2006; Yu & Lui, 2006). This condition implies the negligible duality gap. In other words, the primal problems (3.8) and (3.10) and their associated dual problems (defined later) have almost the same optimal value.

3.3.1 BRA-SPM

Let us denote $\mu_{k,m}$ as Lagrange dual variable for queue m of user k and let overall Lagrange dual variable be a matrix $\boldsymbol{\mu}$ with $[\boldsymbol{\mu}]_{k,m} = \mu_{k,m}$. Then, the Lagrange function for the primal problem (3.8) is defined as

$$\mathcal{L}(\boldsymbol{\rho}, \mathbf{s}, \boldsymbol{\mu}) = \sum_{k=1}^{K} \sum_{m=1}^{M_k} \sum_{n=1}^{N} \rho_{k,m,n} s_{k,m,n} - \sum_{k=1}^{K} \sum_{m=1}^{M_k} \mu_{k,m} \left(\sum_{n=1}^{N} \rho_{k,m,n} r_{k,m,n} - R_{k,m} \right) \quad (3.12)$$

and its dual function $g(\boldsymbol{\mu})$ is

$$g(\boldsymbol{\mu}) = \min_{\{\boldsymbol{\rho}, \mathbf{s}\} \in \Omega} \mathcal{L}(\boldsymbol{\rho}, \mathbf{s}, \boldsymbol{\mu}) \quad (3.13)$$

where the domain Ω is defined as the set of **ρ** satisfying constraint (3.6) and **s** satisfying constraint (3.2).Then, the dual problem is given by

$$V_{SPM}^{*} = \max_{\boldsymbol{\mu}\geq 0} g(\boldsymbol{\mu}) \triangleq \sum_{n=1}^{N} \bar{g}_n(\boldsymbol{\mu}) + \sum_{k=1}^{K}\sum_{m=1}^{M_k} \mu_{k,m} R_{k,m} \tag{3.14}$$

where $g(\boldsymbol{\mu})$ is decomposed into N independent problems $\bar{g}_n(\boldsymbol{\mu})$ which is defined by

$$\bar{g}_n(\boldsymbol{\mu}) = \min_{\{\boldsymbol{\rho},\, \mathbf{s}\}\in\Omega} \left\{ \sum_{k=1}^{K}\sum_{m=1}^{M_k} \rho_{k,m,n}\left(s_{k,m,n} - \mu_{k,m} r_{k,m,n}\right) \right\} \forall n. \tag{3.15}$$

To solve problem (3.15) individually for $n = 1, \dots, N$ under the integer constraint of $\rho_{k,m,n}$, subcarrier n should be assigned to queue m of user k with the minimum value of $G_{k,m,n}(s_{k,m,n}) \triangleq \left(s_{k,m,n} - \mu_{k,m} r_{k,m,n}\right)$. For this reason, we take the derivative of $\sum_{k=1}^{K}\sum_{m=1}^{M_k} G_{k,m,n}(s_{k,m,n})$ regarding $s_{k,m,n}$ which results in

$$\bar{s}_{k,m,n} = \left[\frac{\tau\mu_{k,m}}{T_{ofdm}\ln 2} - \frac{1}{\alpha_{k,m}\gamma_{k,n}}\right]^{+} \forall k, \forall m, \forall n. \tag{3.16}$$

Then, we derive the subcarrier assignment $\bar{\rho}_{k,m,m}$ as

$$\bar{\rho}_{k,m,n} = \begin{cases} 1, & k,m = \arg\min\limits_{k,m} G_{k,m,n}\left(\bar{s}_{k,m,n}\right) \\ 0, & \text{otherwise.} \end{cases} \quad \forall n. \tag{3.17}$$

Now, we substitute equations (3.16) and (3.17) into equation (3.15) and rewrite equation (3.15) as

$$\bar{g}_n(\boldsymbol{\mu}) = \sum_{k=1}^{K}\sum_{m=1}^{M_k} \bar{\rho}_{k,m,n}\left(\bar{s}_{k,m,n} - \mu_{k,m}\bar{r}_{k,m,n}\right). \quad \forall n \tag{3.18}$$

where $\bar{r}_{k,m,n}$ is obtained using equation (3.3) with $s_{k,m,n} = \bar{\rho}_{k,m,n}\bar{s}_{k,m,n}$. After solving equation (3.18) individually for $n = 1, \dots, N$, $g(\boldsymbol{\mu})$ is computed from equation (3.14).

To find the optimal $\boldsymbol{\mu}^*$ that maximizes equation (3.14) to V_{SPM}^* regardless of the dual function differentiability, the update of $\boldsymbol{\mu}$ can be done using the subgradient method in which the subgradient of $g(\boldsymbol{\mu})$ for queue m of user k is defined as

$$\Delta F_{k,m} = R_{k,m} - \sum_{n=1}^{N} \bar{r}_{k,m,n} \quad \forall k, \forall m. \tag{3.19}$$

where $\bar{\rho}_{k,m,n}$, $\bar{s}_{k,m,n}$ and $\bar{r}_{k,m,n}$ are the optimizing variables in the definition of $g(\boldsymbol{\mu}) = \mathcal{L}(\boldsymbol{\rho}, \mathbf{s}, \boldsymbol{\mu})$. Then, $\boldsymbol{\mu}$ is updated by

$$\mu_{k,m}^{i+1} = \left[\mu_{k,m}^{i} + \delta^i \Delta F_{k,m}^{i}\right]^{+} \forall k, \forall m \tag{3.20}$$

where i is the iteration index, δ^i is a positive step size which follows diminishing step size policy (Bertsekas, 1995), and $\Delta F_{k,m}^{i}$ is the subgradient of $g(\boldsymbol{\mu})$ for queue m of user k at the current iteration i. The same iterative procedure is repeated until the subgradients of all queues converge to zero. Finally, $\boldsymbol{\mu} = \boldsymbol{\mu}^*$allows us to attain near optimal solution $\boldsymbol{\rho}^* \approx \bar{\boldsymbol{\rho}}$ and $\mathbf{s}^* \approx \bar{\mathbf{s}}$ because the strong duality holds where $U_{SPM}^* \approx V_{SPM}^*$. The complete steps are given in algorithm 3.1.

Table 3.3 Algorithm 3.1 BRA-SPM optimal solution.

1: Initialization: Set i =1, $\boldsymbol{\mu}^i > 0$, and choose $0 < \epsilon \ll 1$ as the stopping criterion.

2: **repeat**

3: Set $\rho_{k,m,n} = 0 \ \forall k, \forall m, \forall n$.

4: **for** n=1 to N

5: **for** k=1 to K; **for** m=1 to M_k

7: Compute $\bar{s}_{k,m,n}$ using (3.16).

8: **end for; end for**

10: Set $\bar{\rho}_{k,m,n}$ using (3.17).

11: **end for**

12: i=i+1

13: Update $\mu_{k,m}^{i}$ according to (3.20) $\forall k, \forall m$

14: **until** $\left|\Delta F_{k,m}^{i}\right| < \epsilon \quad \forall k, \forall m$

Let I_1 be the number of iterations needed to update $\mu_{k,m}^{i}$ until $\left|\Delta F_{k,m}^{i}\right| < \epsilon$. Then the overall optimization requires the complexity of $\mathcal{O}(I_1 \sum_{k=1}^{K} M_k' N)$. The average I_1 is around 2×10^3 in our experiment.

Proposition 3.1: With the objective of minimizing the total transmit power in a downlink OFDM system, the subcarrier assignment strategy for transmitting bitstreams with multiple target BERs is that subcarriers with good channel conditions should be assigned to important bitstreams with low target BERs, and vice versa.

Proof*:* For ease of explanation, we restrict our attention to the single user case (i.e., $K = 1$) with M_1 queues. Without loss of generality, we assume that the CNRs for the user across N subcarriers satisfy $\gamma_{1,1} \geq \gamma_{1,\dots} \geq \gamma_{1,N}$. Let $\mu_{1,m}^{i} = \mathrm{Z}$ for $m = 1, \dots, M_1$ at the first iteration (i.e., i =1) where Z is a constant. Since the denominator $\ln\left((5\mathrm{BER}_{1,m})^{-1}\right)$ in equation (3.4) is positive and strictly decreasing in the range of $0 < \mathrm{BER}_{1,m} < 0.0446$, it is easy to see that $\alpha_{1,m}$ increases as $\mathrm{BER}_{1,m}$ increases. For brevity, we label this relationship as $\alpha_{1,m} \propto \mathrm{BER}_{1,m}$. Also, the relationship $G_{1,m,n}(\bar{s}_{k,m,n}) \propto (1/\alpha_{1,m})$ has been found empirically for $\bar{s}_{k,m,n} > 0$ as shown in appendix A.

Using the relationships $\alpha_{1,m} \propto \mathrm{BER}_{1,m}$ and $G_{1,m,n}(\bar{s}_{k,m,n}) \propto (1/\alpha_{1,m})$, equations (3.5) and (3.17) imply that subcarrier n is always allocated to queue M_1 (i.e., $\psi_{1,M_1} = \{1, \dots, N\}$) during the first iteration of algorithm 3.1. In other words, the remaining queues (i.e., $m < M_1$) do not occupy any subcarrier for transmission purpose (i.e., $\bar{\rho}_{1,m,n} = 0$ for $\forall m \neq M_1$ and $\forall n$). To satisfy the constraint (3.7), certain subcarriers must be reallocated to queue $m < M_1$ in the following iteration $i > 1$. From the perspective of queue M_1, if it has to relinquish certain subcarriers while minimizing its total transmit power (i.e., $\sum_{n=1}^{N} \bar{\rho}_{1,M_1,n} \bar{s}_{1,M_1,n}$), those subcarriers will be the first N'_{1,M_1} subcarriers with better channel conditions (i.e., $\psi_{1,M_1} \backslash \{1, \dots, N'_{1,M_1}\}$). This is because given the constant α_{1,M_1} across all N subcarriers, we obtain $s'_{1,M_1,n} \propto \gamma_{1,n}$ from the relationship $\bar{s}_{1,m,n} \propto \alpha_{1,m}\gamma_{1,n}$ in equation (3.16) as $\bar{s}_{1,m,n}$ is monotonically increasing for $\alpha_{1,m}\gamma_{1,n} > 0$. Therefore, it is easy to see that relinquishing subcarrier n instead of subcarrier $n + 1$ will always result in larger power reduction. On the other hand, for queues $m < M_1$ which compete for subcarrier n that will be redistributed, it is obvious that minimum total transmit power will achieve when better subcarriers (e.g., $n = 1$) tends to be allocated to MI queues (e.g., $m = 1$) so that $\alpha_{1,m}\gamma_{1,n}$ and its corresponding $s'_{1,m,n}$ could be minimized. This verifies *Proposition 3.1* which is validated by the results presented in section 3.4.

Proposition 3.1 deserves further attention. So far in the literature, several works such as (Sadr, Anpalagan, & Raahemifar, 2009) rely on the concept of preferably treating important bitstreams with good subcarriers in order to enhance the received multimedia quality, without any mathematical reasoning. Therefore, *Proposition 3.1* offers a fundamental insight into such heuristic approach.

3.3.2 BRA-SRM and BRA-SUM

In a similar fashion, the Lagrange function for the primal problem (3.10) is defined as

$$\mathcal{L}(\boldsymbol{\rho}, \mathbf{s}, \lambda) = \sum_{k=1}^{K} \sum_{m=1}^{M_k} \sum_{n=1}^{N} w_{k,m} \rho_{k,m,n} r_{k,m,n} - \lambda \left(\sum_{k=1}^{K} \sum_{m=1}^{M_k} \sum_{n=1}^{N} \rho_{k,m,n} s_{k,m,n} \leq P_{tot} \right) \tag{3.21}$$

and its dual function $g(\lambda)$ is

$$g(\lambda) = \max_{\{\boldsymbol{\rho}, \mathbf{s}\} \in \Omega} \mathcal{L}(\boldsymbol{\rho}, \mathbf{s}, \lambda) \tag{3.22}$$

where the domain Ω is defined as the set of **ρ** satisfying constraint (3.6) and **s** satisfying constraint (3.2).Then, the dual problem is given by

$$V^*_{SRM/SUM} = \min_{\lambda \geq 0} g(\lambda) \triangleq \sum_{n=1}^{N} \bar{g}_n(\lambda) + \lambda P_{tot} \tag{3.23}$$

where $g(\lambda)$ is decomposed into N independent problems $\bar{g}_n(\lambda)$ which is defined by

$$\bar{g}_n(\lambda) = \max_{\{\boldsymbol{\rho}, \mathbf{s}\} \in \Omega} \left\{ \sum_{k=1}^{K} \sum_{m=1}^{M_k} \rho_{k,m,n} \left(w_{k,m} r_{k,n} - \lambda s_{k,n} \right) \right\} \forall n. \tag{3.24}$$

To solve equation (3.24) individually for $n = 1, \ldots, N$ under the integer constraint of $\rho_{k,m,n}$, the subcarrier n should be assigned to user k with the maximum value of $F_{k,m,n}(s_{k,m,n}) \triangleq \left(w_{k,m} r_{k,m,n} - \lambda s_{k,m,n} \right)$. For this, we take the derivative of $\sum_{k=1}^{K} F_{k,m,n}(s_{k,m,n})$ regarding $s_{k,n}$ which results in

$$\bar{s}_{k,m,n} = \left[\frac{w_{k,m} \tau}{\lambda T_{ofdm} \ln 2} - \frac{1}{\alpha_{k,m} \gamma_{k,n}} \right]^{+} \forall k, \forall m, \forall n, \tag{3.25}$$

where $[x]^+$ is defined as $\max(x, 0)$ due to constraint (3.2). Then, we can derive the subcarrier assignment $\bar{\rho}_{k,m,n}$ as

$$\bar{\rho}_{k,m,n} = \begin{cases} 1, & k,m = \arg\max\limits_{k,m} F_{k,m,n}\left(\bar{s}_{k,m,n}\right) \\ 0, & \text{otherwise.} \end{cases} \forall n. \tag{3.26}$$

Now, we substitute equations (3.25) and (3.26) into equation (3.24) and rewrite equation (3.24) as

$$\bar{g}_n(\lambda) = \sum_{k=1}^{K} \sum_{m=1}^{M_k} \bar{\rho}_{k,m,n}\left(w_{k,m}\bar{r}_{k,m,n} - \lambda \bar{s}_{k,m,n}\right). \quad \forall n \tag{3.27}$$

where $\bar{r}_{k,m,n}$ is obtained using (3.3) with $s_{k,m,n} = \bar{\rho}_{k,m,n}\bar{s}_{k,m,n}$. After solving equation (3.27) individually for $= 1, \ldots, N$, $g(\lambda)$ is computed from equation (3.23). To find the optimal λ^* that minimizes (3.23) to $V^*_{SRM/SUM}$, the update of λ can be done using the bisection method until the sum power converges to P_{tot}. Then, $\lambda = \lambda^*$ allows us to attain optimal solution $\boldsymbol{\rho}^* \approx \bar{\boldsymbol{\rho}}$ and $\mathbf{s}^* \approx \bar{\mathbf{s}}$ because the strong duality holds where $U^*_{SRM/SUM} \approx V^*_{SRM/SUM}$. Finally, if the allocation problem is PF, the average throughput of each user k is updated by

$$\overline{R_{k,m}}[t+1] = (1 - 1/t_c)\overline{R_{k,m}}[t] + (1/t_c)\sum_{n=1}^{N} \rho^*_{k,m,n} r^*_{k,m,n} \quad \forall k, \forall m \tag{3.28}$$

where t_c is a low pass filtering parameter for rate averaging purpose and the second term on the right hand side of equation (3.28) represent the total number of transmit bits allocated to user k at the current time slot t.The complete steps are given in algorithm 3.2.

Table 3.4 Algorithm 3.2 BRA-SRM and BRA-SUM optimal solutions.

1: Initialization: Choose two points a and b such that the optimal point λ^* lies in the interval (a, b). Also choose $0 < \epsilon \ll 1$ as the stopping criterion. Set $\lambda_{low} = a, \lambda_{high} = b$, and $\lambda = (\lambda_{low} + \lambda_{high})/2$. Set current time slot $t = 1$.

2: Choose BRA scheme from {BRA-SRM ,BRA-SUM (PF)}.

3: If BRA=MTC, set $w_{k,m} = 1\ \forall k$.

4: If BRA=PF AND $t = 1$, set $w_{k,m} = 1$ and $\overline{R_{k,m}}[t] = 0\ \forall k$. Else if If BRA=PF AND $t \neq 1$, set $w_{k,m}$ using (3.11).

5: **repeat**

6: Set $\rho_{k,m,n} = 0\ \forall k, \forall m, \forall n$.

7: **for** n=1 to N

8: **for** k=1 to K; **for** m=1 to M_k

9: Compute $\bar{s}_{k,m,n}$ using (3.25).

10: **end for; end for**

11: Set $\bar{\rho}_{k,m,n}$ using (3.26).

12: **end for**

13: If $\sum_{k=1}^{K}\sum_{m=1}^{M_k}\sum_{n=1}^{N}\bar{\rho}_{k,m,n}\bar{s}_{k,m,n} > P_{tot}$, set $\lambda_{low} = \lambda$; Else, set $\lambda_{high} = \lambda$.

14: **until** $\left|\sum_{k=1}^{K}\sum_{m=1}^{M_k}\sum_{n=1}^{N}\bar{\rho}_{k,m,n}\bar{s}_{k,m,n} - P_{tot}\right| < \epsilon$

15: If BRA=PF, update $\overline{R_{k,m}}[t+1]$using (3.28).

16: $t = t + 1$

3.4 Simulation Results

This section is divided into two parts. Each part compares the performance of BRA and WBRA schemes. For WBRA schemes, they treat all bitstreams equally. Therefore, the selected BER constraint has to be the strictest target BER among multiple target BERs.

3.4.1 BRA-SPM

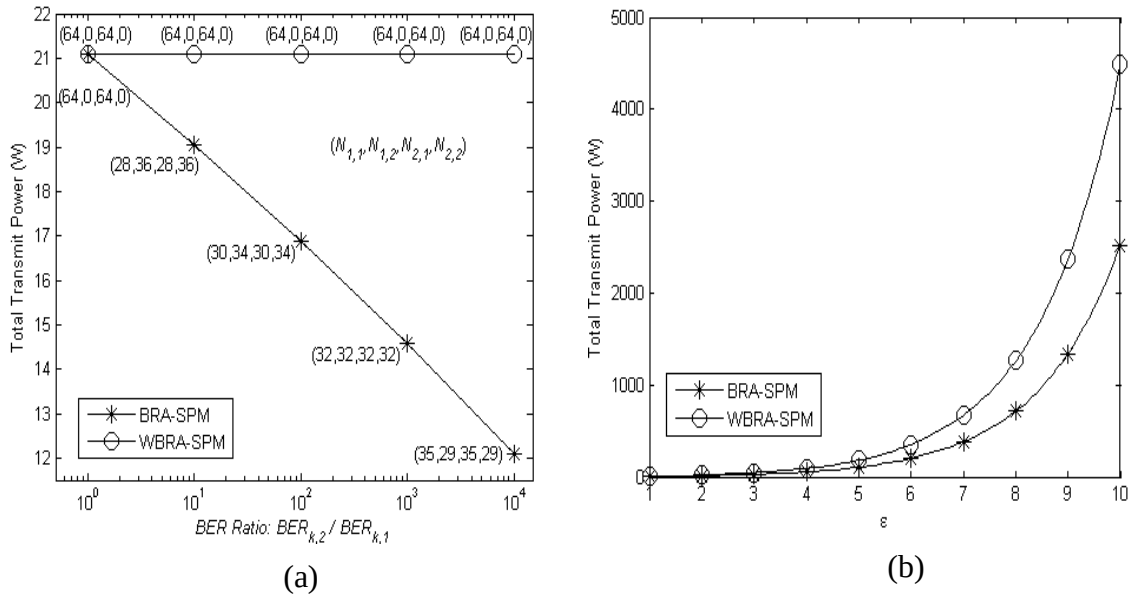
To verify the robustness of BRA-SPM scheme, two sets of simulation parameters were considered. For the first set of parameters, the time slot duration $\tau = T_{ofdm}$, the number of subcarriers $N = 128$, the number of users $K = 2$, the number of queues for each user $M_1 = M_2 = 2$, $\mathrm{BER}_{k,1} = 10^{-6}$, and the rate constraints for each queue $R_{k,m} = 50$ bits for $\forall k$ and $\forall m$. The first and second row vectors of CNR matrix $\boldsymbol{\gamma}$ were set to $[1:128]$ and $[128:1]$, respectively. According to the above parameters and *proposition 1*, $\psi_{1,1} \cup \psi_{1,2} \cup \psi_{2,1} \cup \psi_{2,2} = \{1, \dots, 128\}$, $|\psi_{1,1}| = N_{1,1}$, $|\psi_{1,2}| = N_{1,2}$, $|\psi_{2,1}| = N_{2,1}$, $|\psi_{2,2}| = N_{2,2}$ and

$$\underset{\psi_{2,1}\ \ \psi_{2,2}\ \ \psi_{1,2}\ \ \psi_{1,1}}{\{1, \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots, 128\}}, \tag{3.29}$$

Note that the first simulation intends to illustrate *proposition 3.1* with simple analysis. Let ($N_{1,1}$, $N_{1,2}$, $N_{2,1}$, $N_{2,2}$) be the vector of number of assigned subcarriers. $N_{1,2}$ and $N_{2,2}$ are always zero for WBRA case since each user only has single queue. In other words, (64,0,64,0) indicates that 64 subcarriers are adopted to transmit all incoming packets for each user. For the second set, $\tau = T_{ofdm}$, $N = 256$, $K = 4$, $M_k = 3$, $BER_{k,1} = 10^{-6}$, $BER_{k,2} = 10^{-4}$,and $BER_{k,3} = 10^{-2}$ for $\forall k$ and $\forall m$. The first, second, third and fourth row vectors of CNR matrix $\boldsymbol{\gamma}$ were set to $[1:256]$, $[256:1]$, $2 * [1:256]$,

$2 * [256: 1]$, respectively. Both first and second simulation results are plotted in figure 3.2 (a) and (b), respectively.

Figure 3.2 (a) shows the total transmit power of BRA-SPM and WBRA-SPM schemes under different target BERs of queue $m = 2$, i.e., $\text{BER}_{1,2}$=$\text{BER}_{2,2} = 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}$ and 10^{-6}. Hence, the BER ratio $\text{BER}_{k,2}/\text{BER}_{k,1}$ takes value from 1 to 10^4. As expected, the total transmit power for WBRA-SPM remains flat since it assumes no discrepancy in importance between the MI and the LI bitstreams. Whereas for BRA-SPM, the total transmit power and the number of subcarriers assigned to the LI bitstreams decreases when the BER ratio increases. This is because when loosing the BER constraints of the LI bitstreams, more increase of transmit rate can be achieved by one unit of power consumption according to equation (3.3). Hence, less power and corresponding less subcarriers are required to empty these LI buffers on the one hand, and more subcarriers and corresponding less power are needed to transmit the MI bitstreams on the other hand. Figure 3.2 (b) depicts the total



(a) (b)

Figure 3.2 Total transmit power under different constraints. (a) BER ratio for different value of $\text{BER}_{k,2}$ with a fixed data rate and $\text{BER}_{k,1}$ in $K = 2, M_k = 2$ case. (b) Data rate $20 * \varepsilon$ bits for all queues with a fixed $\text{BER}_{k,m}$ in $K = 4, M_k = 3$ case.

transmit power of BRA-SPM and WBRA-SPM schemes for different rate constraints, i.e., $R_{k,m} = 20 * \varepsilon$ bits for $\forall k$ and $\forall m$. It can be observed that the total transmit power for both methods increases with the rate constraints. However, BRA-SPM significantly outperforms WBRA-SPM where the performance gap widens as the rate constraints increase. This is because when the size of data in the LI bitstreams (i.e., $m = 2,3$) increases, more power reduction can be achieved by BRA-SPM than WBRA-SPM.

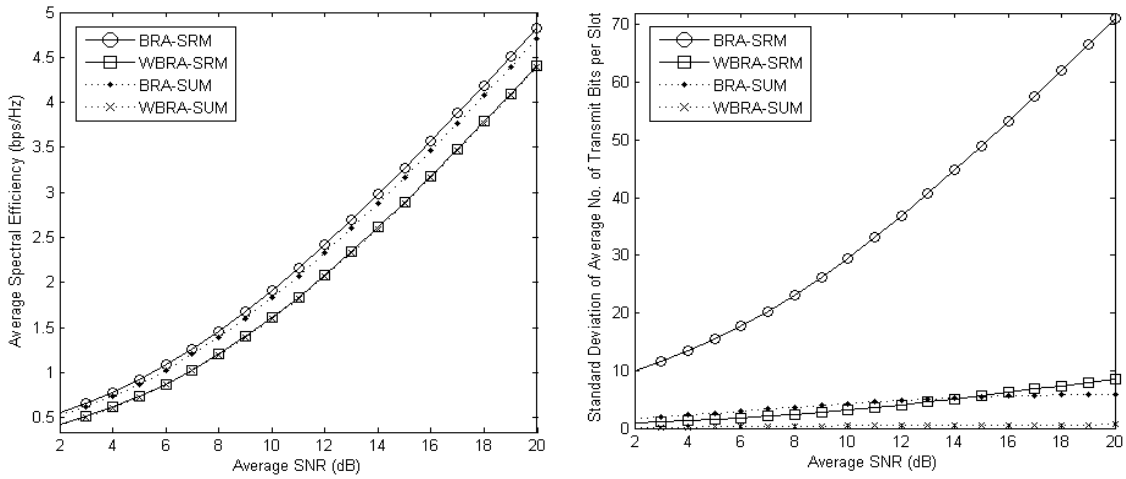
3.4.2 BRA-SRM and BRA-SUM

For the OFDMA system, the total bandwidth $B_T = 1$ MHz, the number of subcarriers $N = 128$, the number of users $K = 6$, the number of queues for each user $M_k = 1$, and the time slot duration $\tau = T_{ofdm}$. The wireless channel was modeled as a frequency-selective channel that consists of six independent Rayleigh multipaths. Each multipath component was modeled by Clarke's flat fading model (Rappaport, 2002). We assume the power delay profile exponentially decays with e^{-2v}, where v is the multipath index. Then, the relative power of the six multipath components are [0, −8.69, −17.37, −26.06, −34.74, −43.43] dB. We assume there is no path loss difference between the users, thus the average channel gain $E\left[\left|H_{k,n}\right|^2\right]$ for all k and n is set to one. The maximum Doppler shift is 30 Hz. The power spectral density of the white Gaussian noise is -80 dBW/Hz. The average signal-to-noise ratio (SNR) is defined as $(P_{tot}/B_T N_0)$. All the results were averaged over 1000 time slots. We define average spectral efficiency as

$$\text{average spectral efficiency} = \left(\sum_{k=1}^{K} \sum_{m=1}^{\boldsymbol{M_k}} \sum_{n=1}^{N} \rho_{k,m,n}^{*} r_{k,m,n}^{*} \right) / N \qquad (3.30)$$

For the PF-like schemes, we set t_c=500 (Sadiq, Madan, & Sampath, 2009).

Figure 3.3 shows the average spectral efficiency and the standard deviation of the average number of transmit bits per time slot across the users versus the average SNR



(a) (b)

Figure 3.3 Network capacity versus average SNR over all users.
(a) Average spectral efficiency. (b) Standard deviation of the average number of transmit bits per time slot.

for different methods. Note that the smaller the value of standard deviation, the better the user fairness. We set the target BER of each user as follows: $\mathrm{BER}_{1,1} = 10^{-6}, \mathrm{BER}_{2,1} = 10^{-6}, \mathrm{BER}_{3,1} = 10^{-5}, \mathrm{BER}_{4,1} = 10^{-5}, \mathrm{BER}_{5,1} = 10^{-4}$ and $\mathrm{BER}_{6,1} = 10^{-4}$. In figure 3.3 (a), it can be seen that the average spectral efficiency for all methods increases with the average SNR and the BRA-like schemes outperform the WBRA-like schemes for all the average SNR values. The performance gaps between the BRA-like schemes and the WBRA-like schemes increase as the average SNR increases. This discrepancy is as high as 0.4113 bps/Hz for the SRM case and 0.3166 bps/Hz for the PF case. This is because with a loose target BER, more increase of transmit rate can be achieved by one unit of transmit power according to equation (3.3). Hence, as P_{tot} increases with the average SNR, more increase of spectral efficiency can be achieved by the BRA-like schemes.

Besides that, SRM-like schemes achieves higher average spectral efficiency than PF-like schemes as expected. However, the average spectral efficiency for WBRA-SRM scheme is just marginally higher than that for WBRA-PF. This is surprising as the

SRM-like schemes solely maximize the average spectral efficiency while the PF-like schemes enforce fairness among users by trading off the spectral efficiency. The reason is that where there is no path loss difference between the users, the spectral efficiency or equivalently the sum capacity is not very sensitive to the fairness issue (Shen et al., 2005). In contrast, a wider gap in the average spectral efficiency exists between the BRA-SRM and BRA-PF schemes. Again, this is due to loose BER constraints.

In figure 3.3 (b), we can observe that the WBRA-SRM, BRA-PF and WBRA-PF schemes tend to provide similar throughput across all users and thus exhibit low standard deviation. Among these three methods, the WBRA-PF scheme has the most uniform throughput across all users due to only one BER constraint and its optimization objective. The BRA-SRM scheme incurs much larger standard deviation than other alternatives, because it always serves the user with the highest target BER and the best channel condition in priority.

Figure 3.4 depicts the average spectral efficiency for different schemes for different target BER of user $k = 6$, i.e., $\mathrm{BER}_{6,1} = 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}\ and\ 10^{-6}$ while fixing $\mathrm{BER}_{1,1} = \mathrm{BER}_{2,1} = \mathrm{BER}_{3,1} = \mathrm{BER}_{4,1} = \mathrm{BER}_{5,1} = 10^{-6}$. Hence, the BER ratio $\mathrm{BER}_{6,1}/BER_{k\neq 6,1}$ takes value from 1 to 10^4. It can be seen that the BRA-like and WBRA-like schemes exhibit same performance when the BER ratio is one. However, when the BER ratio increases, the average spectral efficiency for the BRA-like schemes increases drastically, whereas the average spectral efficiency for the WBRA-like schemes is flat. The reason is that $\mathrm{BER}_{6,1}$ increases when the BER ratio increases. Accordingly, more increase of transmit rate can be achieved by one unit of transmit power when loosing BER constraint. In contrast, for the WBRA-like schemes, the BER constraint is determined only by the strictest target BER, which is 10^{-6}. Besides that,

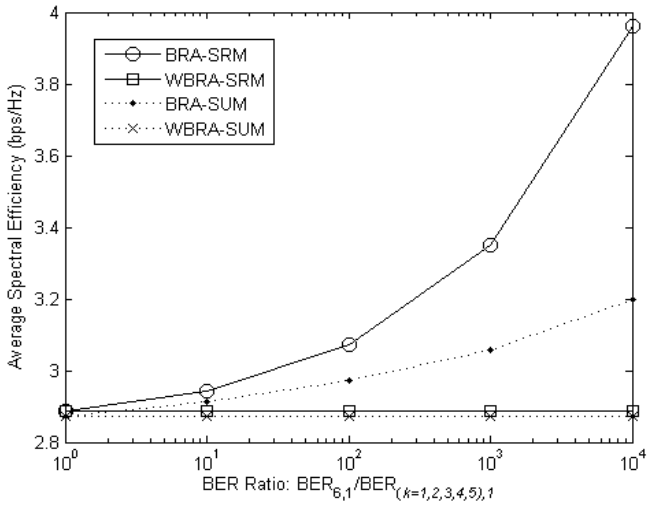


Figure 3.4 Average Spectral Efficiency versus BER ratio for different. value of $BER_{k,6}$.

the average spectral efficiency for the SRM-like schemes is always higher than that for the PF-like schemes as expected.

3.5 Summary

In this chapter, we have developed three BER-driven resource allocation schemes namely BRA-SPM, BRA-SRM and BRA-SUM, corresponding to three general types of resource allocation. An important theorem that important multimedia bitstreams should be assigned subcarriers with good channel conditions has been proved. We demonstrated the advantages of BRA schemes over WBRA schemes in terms of spectral efficiency and power efficiency. Although higher resource utilization is obtained, this may not reflect the actual received multimedia quality since the performance gain is attained at the cost of lower overall transmission reliability. For this reason, we next turn to the problem of BRA for realistic multimedia bitstreams.

CHAPTER 4

BER-Driven Resource Allocation (BRA) for Realistic Nonscalable Multimedia Bitstreams

Previous chapter boosts the overall OFDMA performance by leveraging the characteristic of multiple target BERs residing in generic bitstreams. Such attribute also holds for a nonscalable video bitstream as different video segments possess different levels of importance in terms of reconstructed quality. In this chapter, the challenge is further magnified when adaptive target BERs, real source coding, practical modulation and coding schemes, as well as source-level error resilient methods are taken into account. This chapter ends with a simulation study detailing the performance of the proposed scheme in terms of received quality.

4.1 Introduction

OFDMA system, thanks to its superior spectral efficiency, is the right candidate for accommodating the high-speed demand of video communications. However, wireless networks generally cannot guarantee error-free delivery of data due to hostile channel conditions. To add insult to injury, retransmission of corrupted video packets may not be feasible due to stringent delay constraint. A single error in a video packet may induce severe error propagation in spatial-temporal domains.

To achieve robust video transmission over noisy channels, popular video coding standards such as H.264/AVC ("ISO/IEC 14496-10/2005," 2007) provide several decoder-based error resilient tools to compensate the effect of transmission errors. However, prevalent approaches assume a packet-loss scenario, where the decoder discards and conceals all the compressed video information within an erroneous packet, even if only a few bit is flipped (Larzon, Degermark, & Pink, 1999; Wenger, 2003). Indeed, the damaged packet still contains redundancy that can be exploited to determine

more exactly the error position, and to possibly recover a part of it (Superiori, Nemethova, & Rupp, 2006), especially the head portion preceding the first bit error. This property has been frequently utilized to optimize the received video quality (Caron & Coulombe, 2012; Farrugia & Debono, 2010; Ko, Hong, & Suh, 2012).

Although decoder-based error resilient technique is a powerful tool to achieve robustness against packet error, it cannot always overcome the abrupt channel conditions induced by the frequency-selective time-varying fading. This problem becomes especially relevant in OFDMA networks, due to the frequency-selectivity that results in different fading levels experienced by different subcarriers. Coded OFDM system is an efficient means to address such issue with channel coding and interleaving (Lefloch, Alard, & Berrou, 1995). In particular, the input data are encoded into coded symbols that will be interleaved across several subcarriers. In this way, the errors caused by poor subcarriers can be possibly corrected using the reliable data of good subcarriers. Starting from these premises, it is clear that resource allocation, error resilience and coded OFDM are important for robust data transmission. Thus, this chapter focuses on the interaction among these three aspects in the context of video communications.

Recently, enormous research has been devoted to video communications over OFDMA systems. One of the active research areas is the multiuser pre-encoded video streaming (Ji et al., 2009; Karray, 2010; K. D. Lee & Leung, 2009; F. Li et al., 2012; P. Li, Chang, Feng, & Yang, 2011; Mokari, Javan, & Navaie, 2010; Pahalawatta et al., 2007), where the focus has been on how the video quality is improved using effective packet scheduling and resource allocation approaches.

In (Pahalawatta et al., 2007), a scheme that maximizes the weighted sum rate of all video users is proposed. Each user weight is measured by the highest importance level

of packet found in that user transmission queue. Similar strategy is advocated in (Ji et al., 2009) for the case of scalable video, except for the additional parameter called packet deadline requirement being included in the weight calculation. The above two schemes, however, share the same drawback, i.e., the users with much higher weights always seize the transmission opportunity. This may lead to resource starvation in the remaining users. Such unfairness issue is addressed in (P. Li et al., 2011) by taking into account the scheduled packet size. However, identical to (Ji et al., 2009; Pahalawatta et al., 2007), the packets are scheduled first by neglecting their resource consumption efficiency. A step forward from allocating resources for a group of scheduled packets is the joint consideration of these two aspects. The authors in (F. Li et al., 2012) demonstrate the promising video performance with such design.

Despite the significant gains accrued by the above-mentioned works, their video performances are purely derived based on the assumption of error-free transmission. This, however, do not hold in practical communication systems that involve modulation and coding techniques. In fact, in such systems, the prime concern is to attain the best tradeoff between the bit rate and the transmission reliability associated with a target bit-error-rate (BER). The target BER criterion has a critical impact on the video performance (D. W. Wang, Toni, Cosman, & Milstein, 2013). If the value is loose, despite the abundant availability of packets at the receiver, most of them are corrupted and thus undecodable. Instead, if the target BER is strict, packets are forwarded in a conservative way, which may render limited achievable quality though most of the packets are decodable.

In (Yang & Wang, 2013), practical modulation and coding schemes are considered in their video transmission framework. However, similar to (D. W. Wang et al., 2013), all video packets are simply assigned with a static target BER, irrespective of the

importance of that packet; moreover, equal power allocation (EPA) is assumed in which a total power budget is distributed evenly over all subcarriers.

In this chapter, a three-level BER-driven video transmission framework is proposed to optimize the packet scheduling, subcarrier allocation as well as bit and power allocation. The ultimate goal is to maximize the average video quality of all users under the playback delay and resource constraints. This is achieved by jointly exploiting the multiuser diversity and the video characteristics. The main contributions of this chapter are as follows.

First, most existing works treat all the video packets with the same target BER. In contrast, the proposed method entails that more important video packets are assigned with stricter target BERs, and *vice versa*. This can be viewed as a form of unequal error protection (UEP) which is established for its superior performance in transmission over noisy networks. Second, an *overall target BER* is considered to control how many average packets can be sent within one time slot. As observed in (D. W. Wang et al., 2013), the transmitter itself does not know the optimal value that renders the best overall video quality. In this regard, the proposed method seeks the best quality-resource tradeoff and thus will always be a safer option irrespective of the pre-determined *overall target BER*.

Third, with a practical wireless video transmission system, this chapter aims to cast some light on the suitability of EPA, which is the most prevalently used alleviation in the resource-allocation literature (H. W. Lee & Chong, 2008). In fact, most previous works justify the use of EPA from a theoretical perspective and isolate the influence of channel coding. In particular, validations are made by distinguishing the performance between EPA and frequency-selective power allocation (FPA) in terms of spectral efficiency, where FPA assigns power to subcarriers as a function of their channel

conditions. It is however, not clear, how exactly the theoretical performance gap translates into the video performance gap. The findings are deemed valuable as channel coding is part and parcel of practical communication systems.

The rest of this chapter is organized as follows. Section 4.2 provides an overview of H.264/AVC coding. Section 4.3 describes our system model. In section 4.4, we elucidate how to estimate the importance level of each packet. Then, we propose a three-level framework consisting of packet scheduling, subcarrier assignment and bit and power allocation. Section 4.5 provides simulation details and results. We wrap up the chapter in section 4.6. A list of the key mathematical notations used in this chapter is provided in table 4.1.

4.2 Overview of H.264/AVC

H.264/AVC (for short, H.264) is a widely accepted video coding standard jointly developed by the ISO/IEC Motion Picture Experts Group (MPEG) and ITU-T Video

Table 4.1 List of Key Notations

Notation	Description
overall target BER	Control how many average packets that can be transmitted within one scheduling slot
MB	macroblock
$R_{k,m}$	Length of video packet m of user k
$Q_{k,m}$	Contribution of packet m of user k to the reconstructed quality
$T_{k,m}$	Waiting time of packet m of user k
T_{max}	Playback deadline
f	Frame f
$\omega_{k,m}(f-1)$	Number of successfully decoded MBs for each transmitted packet m of user k for frame $f-1$
$\overline{\omega}$	Set of overall number of correctly decoded MBs for each

	transmitted packet up to frame $f-1$
M_k	Number of available video packets per frame
$\mathcal{B}/\mathbb{C}$	Coding rate
$s_{k,n}$	Power allocated on subcarrier n for user k
$r_{k,n}$	Number of transmit bits of subcarrier n for user k
$\alpha_{k,n}$	Coding loss factor related to $\mathrm{BER}_{k,n}$ by adopting MQAM modulation
$\Pi_k(f, m-1)$	Reordered set of packets in queue k
$D^L_{k,f,m}$	Estimated distortion caused by the lost of packet m after the transmission of $\Pi_k(f, m-1)$ with $\bar{\omega}$
$D^T_{k,f,m}$	Estimated distortion caused by the transmission of packet m after the transmission of $\Pi_k(f, m-1)$ with $\bar{\omega}$
x	width of the video frame
y	height of the video frame
$D^L_{k,f,m,mb}$	Estimated distortion caused by the lost of MB mb of packet m of frame f of user k
$D^T_{k,f,m,mb}$	Estimated distortion caused by the transmission of MB mb of packet m of frame f of user k
W_{intra}	Weight factor of intra MB
W_{inter}	Weight factor of inter MB
$MV^x_{mb,j}$	Motion vector in x axis
$MV^y_{mb,j}$	Motion vector in y axis
X, Y	Partition mode in which the H.264 standard supports seven MB partition modes
A_m	Area of packet m
A_1	Area of overlapped regions between the current MB and the reference MBs in packet m' of the previous frame
A_2	Area of overlapped regions between the current MB and the reference MBs in packet m'' of the previous frame
$MB_{k,f,m}$	number of MBs in packet m of frame f of user k
$P_{k,f,m,mb}$	Probability that a bit error occurs in MB mb
$R_{k,f,m,mb}$	Length of MB mb
W_k	Weight associated with the importance level of scheduled

	video packets
$\Gamma_{k,n}$	Modulation level on subcarrier n for user k
$b_{k,n}$	Number of bits on subcarrier n for user k
b_{max}	Maximum number of bits
P_{tot}	Total available power
Θ	Constraint length of convolutional coding
$J_{k,m}$	Number of remaining subcarriers

Coding Experts Group (VCEG). Some features in H.264 especially those important to this dissertation (from the communication perspective) are highlighted below. A detailed overview of the entire H.264 codec is available in (Richardson, 2010; Wenger, 2003).

4.2.1 H.264 Syntax

H.264 compresses a video sequence that is composed of several frames in time order as shown in figure 4.1. Each frame (picture) is further split into non-overlapping pixel-areas called macroblocks (MBs) consisting of 16x16 samples of the luma and 8x8 samples of each of the two chroma components. The MBs are grouped into slices, each representing subsets of MBs that can be decoded independently, with the slice header acting as a resynchronization marker. In other words, the slices can be transmitted out of order and still be decoded correctly at the decoder. This crucial feature known as **Arbitrary Slice Ordering** has propelled many research efforts on content-aware scheduling approaches where more important video packets are given higher priorities.

Generally speaking, the importance of each video packet is very much determined by the type of slice it belongs to. There are three types of slices: intra (I), predictive (P), and bi-predictive (B). An I slice contains only I MBs, a P slice may contain both I and P MBs and a B slice may consist of I and B MBs. An I MB is predicted based on spatially

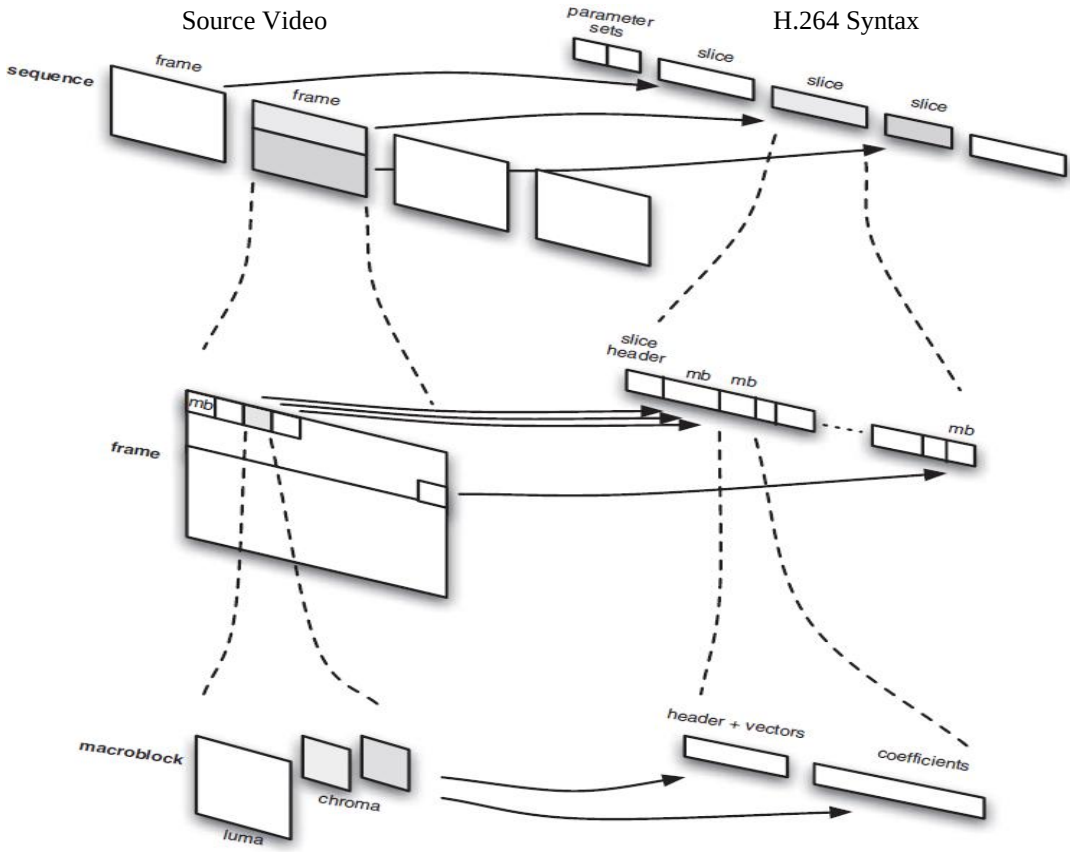


Figure 4.1 H.264 syntax: high-level overview (Richardson, 2010).

adjacent samples of previously coded MBs within the same slice and it is referred to as intra (spatial) prediction. Typically, all the MBs for the first frame of a sequence adopt such prediction. There are several prediction modes available for intra prediction and a subset of these modes is available for various MB partition modes of block sizes, ranging from 16x16 luma, through 4x4 luma, to 8x8 chroma blocks. Among all these modes, the MB directly above the current I MB and the one directly to the left are often employed to extrapolate the estimated pixel values of the current I MB. Then, only the residual block is transformed via an integer 4x4 discrete cosine transform (DCT), quantized using a uniform quantizer and entropy coded.

A P MB is predicted using temporally neighboring MBs of previously coded pictures and it is referred to as inter (temporal) prediction. First, motion estimation for each MB is performed by finding the best matching region from reference pictures. Each MB can be partitioned into smaller blocks for the matching purpose in order to yield the least

total estimation error. H.264 supports seven partition modes of block sizes 16x16, 16x8, 8x16, 8x8, 8x4, 4x8, and 4x4. Second, motion compensation is applied by subtracting the chosen candidate region from the current MB to form a residual block. Third, the residual block is encoded and transmitted alongside a motion vector (MV) indicating the position of the selected region relative to the current MB position. For a B MB, the operations are analogous to those of P MBs except that the reference frames are not only previous frames but also can be successive one. Similar to H.264 intra coding, the residual block goes through transform, quantization and entropy coding.

To reconstruct a MB at the receiver, the transmitted residual needs to be added to the predicted signal. As I MBs generally contain more video information than P and B MBs, losing them would culminate in severe spatial and temporal error propagation effects. This asserts those I packets corresponding to I MBs as high-priority packets during the transmission across networks.

4.2.2 Transport Mechanisms

While H.264 precisely defines syntax for representing compressed video and related information, it does not impose the adoption of any specific transmission mechanism for these data. Rather, it leaves vendors free to implement their own solution, in a way most appropriate for a particular network such as IP-based networks. Such flexibility is guaranteed by an interface layer called Network Abstraction Layer (NAL). Within this layer, all video data are encapsulated in NAL Units (NALU), which can be divided into Video Coding Layer (VCL) NALUs and non-VCL NALUs as illustrated in figure 4.2.

VCL NALUs contain the compressed video data such as encoded MVs and encoded quantized DCT coefficients. Non-VCL NALUs contain control parameters such as Sequence Parameter Sets (SPS) and Picture Parameter Sets (PPS). SPS carry decoding parameters common to an entire video sequence while PPS contain parameters common

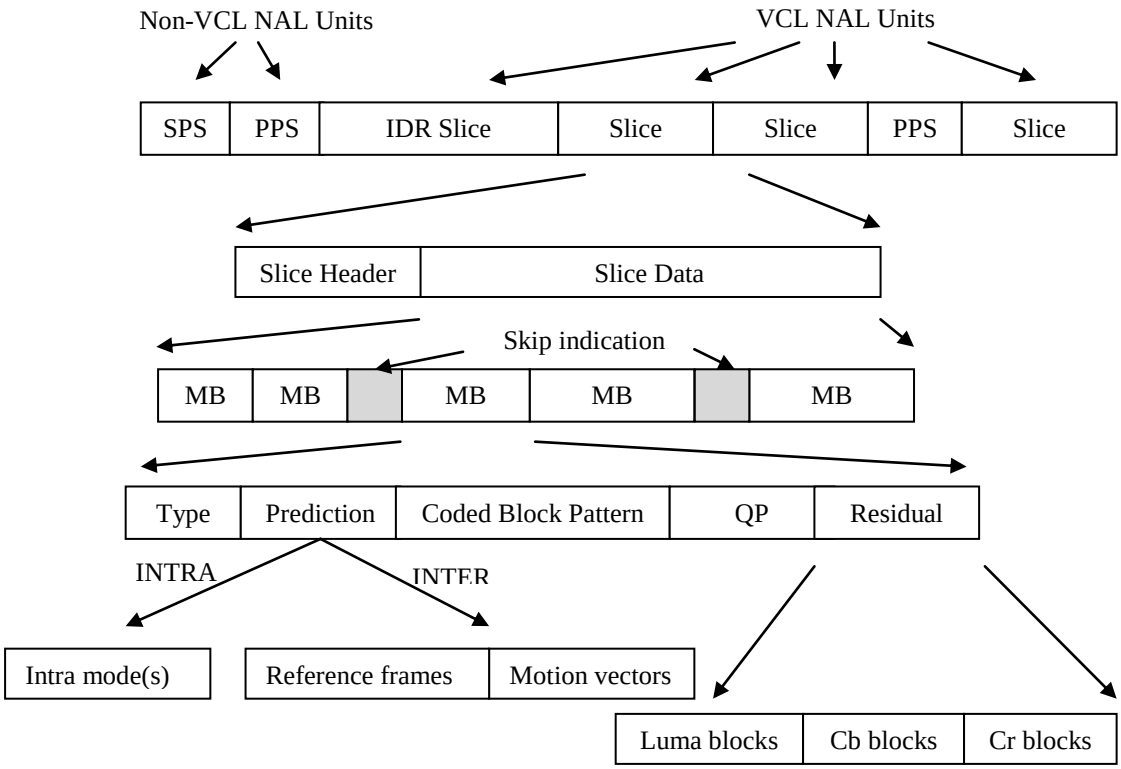


Figure 4.2 H.264 syntax: low-level overview (Richardson, 2010)

to all slices in a single frame. Note that SPS and PPS are encapsulated in separate NALUs. Similarly, a VCL NALU contains only a single slice. Since almost all modern networks including 4G are operating in IP-packet mode, it is naturally to packetize these NALUs into a series of real-time transport protocol (RTP) or user datagram protocol (UDP) packets.

4.2.3 Error Resilience Features

To protect the compressed bitstream in an error-prone transport environment, H.264 has standardized several error resilience features as follows. A broad review of error resilience techniques is available in (Kumar, Xu, Mandal, & Panchanathan, 2006).

Intra refreshing limits the temporal error propagation by inserting I MBs in the bitstream at intervals.

Flexible Slice Size allows the slices to contain any number of MBs in a frame. On the one hand, larger slice sizes lead to higher compression efficiency but come at the cost of

lower degree of error tolerance. On the other hand, smaller slice sizes exhibit greater resilience to packet loss (Connie et al., 2008).

Constrained Intra Prediction forbids previously adjacent coded MBs whose coding mode is either P or B to become reference MBs. Instead, only pixels from other I MBs can be used for intra prediction. This argument is based on the fact that inter MBs could easily lead to spatial error propagation as the MBs themselves rely on other MBs for an accurate reconstruction.

All the above mentioned features can be categorized as encoder-based error resilient technique as they are enabled prior to a transmission. While these techniques enhance the robustness against data loss, they share the same drawback, i.e., additional coding redundancy is necessary.

Another category called decoder-based error resilient strategies does not introduce such redundancy. Error concealment, under this category, is part and parcel of H.264 although it has been extensively used in prior video standards. The key principle of video-based error concealment is to exploit spatial and temporal correlations in the video data so that missing pixels are estimated using data from the received slices.

One straightforward solution of spatial based error concealment is **Weighted Pixel Value Averaging** where the weight depends on the distance from the concealed pixels. That is, each pixel of a missing MB is interpolated as a linear combination of the nearest pixels in the boundaries. For temporal domain based error concealment, the simplest method is **Temporal Replacement** (TR) method. In this method, each pixel value of the concealed slice is directly copied from the corresponding pixel of the last decoded frame. More complex temporal based concealment methods do exist. For example, the MVs of adjacent decoded MB’s in the same frame can be used to predict the MV of a

lost MB. Note that encoder-based and decoder-based error resilient strategies can be combined to yield complementary benefits.

Apart from the techniques standardized by H.264, one specific on-going research line that is worth looking at is the partial slice decoding (PSD) (Demirtas, Reibman, & Jafarkhani, 2011; Trudeau, Coulombe, & Pigeon, 2011). The motivation behind this evolving technique is that the information contained within a corrupted slice is indeed utilisable (Kambhatla, Kumar, Paluri, & Cosman, 2012; Larzon et al., 1999). Substantial video quality improvement can be extracted from those damaged slices (Caron & Coulombe, 2012). H.264, however, assumes a packet-loss scenario, where the receiver simply discards and conceals an entire slice whenever an error is detected.

One possible PSD approach is to apply H.264 syntax verification rules. Once a syntax-decoding violation arises, the MBs preceding the error are retained while the potentially affected and the rest MBs are concealed with an existing concealment method (Superiori et al., 2006). An advanced technique is to apply joint source-channel decoding (JSCD) technique in order to potentially correct the corrupted slices (Caron & Coulombe, 2012; Farrugia & Debono, 2011; Levine, Lynch, & Le-Ngoc, 2007; Nguyen, Lynch, & Tho, 2010; Y. Wang & Yu, 2005). Although there are some undetectable or unrecoverable errors, all the above PSD works demonstrate the promising performance of single-user video transmission with peak signal-to-noise ratio (PSNR) gains of up to 18.1 dB. Distinguishable from these studies, we focus on how resources can be allocated to fully utilize the strength of PSD in the context of multiuser video transmission.

4.3 System Overview

Figure 4.3 depicts the three-level framework, where a BS transmits different video streams to K users. The BS obtains the pre-encoded H.264 video streams from a remote

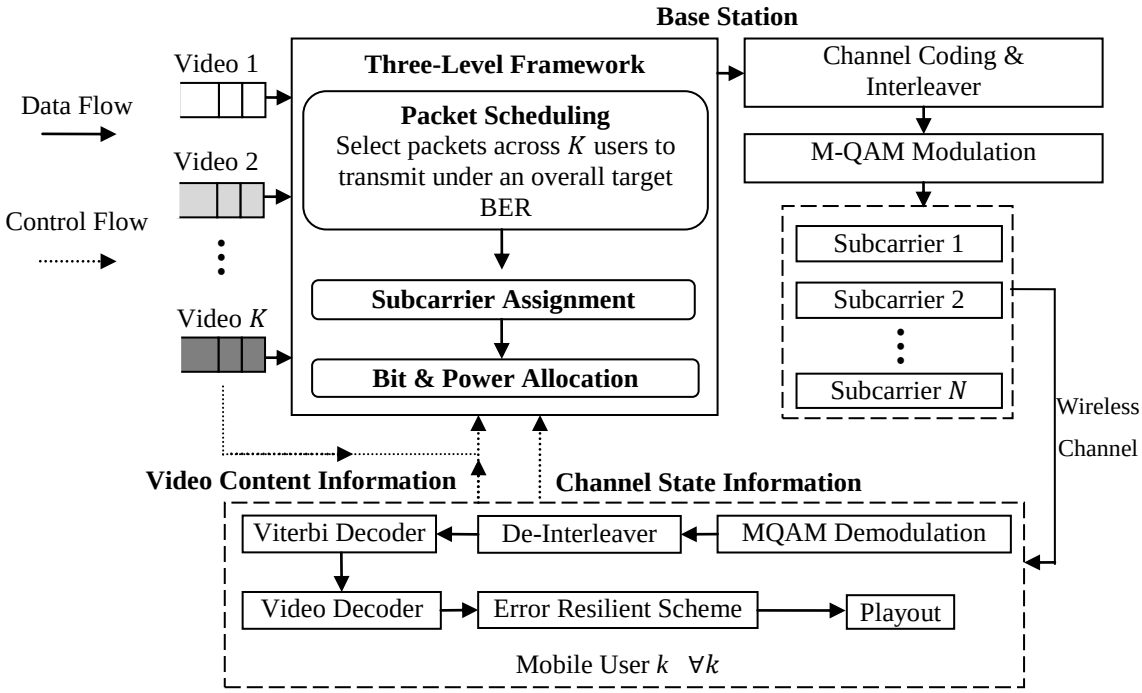


Figure 4.3 System block diagram for downlink OFDMA transmission of video bitstreams.

media server via a lossless network and buffers these streams in different queues. We assume that in one video frame, each slice consists of a row of MBs. The slice is directly packetized into a transport packet where the length of slice (packet) m of user k (in bits) is denoted by $R_{k,m}$. Hereafter the term "slice" and "packet" will be used interchangeably. Different packets from different users entail different contributions to the reconstructed quality $Q_{k,m}$ (defined in section 4.4.1). Additionally, each packet is stamped with a waiting time $T_{k,m}$. Any packet that misses its playback deadline T_{max} will be dropped at the BS. In this case, the number of time slots available per video frame can be calculated by $\lfloor T_{max}/\tau \rfloor$.

We consider two decoder-based error resilient approaches that work perfectly: entire (simple) slice decoding (ESD) and partial (complex) slice decoding (PSD). The former discards and conceals an entire slice regardless of the bit error position. For the latter, when an intermediate MB is in error, the error-free MBs of the slice preceding the corrupted MB are correctly decoded, and the remaining MBs are concealed by an error

concealment strategy. Note that perfect error detection in PSD case is an ideal assumption as it is rarely possible to detect every syntax error. However, as discussed earlier, the superiority of PSD over ESD is still impressive even in the presence of undetectable errors. Once every previous frame $f-1$ is decoded, the number of successfully decoded MBs for each transmitted packet m of user k $\omega_{k,m}(f-1)$ is sent back to the BS via an error-free feedback channel. For notational convenience, let the overall number of correctly decoded MBs for each transmitted packet as a matrix $\boldsymbol{\omega}$ with $[\boldsymbol{\omega}]_{k,m} = \omega_{k,m}$. The BS will update this information from the first frame to the previous frame $f-1$ in the set $\bar{\omega} = \{\boldsymbol{\omega}(1), \boldsymbol{\omega}(2), \dots, \boldsymbol{\omega}(f-1)\}$.

All scheduled video sessions are transmitted via an OFDMA system that is based on the bit-interleaved coded-modulation (BICM) as explained in section 2.5.1. Unless specified otherwise, the assumptions, notations, and parameters of the OFDMA system are the same as in chapter 3. First, equations (3.2)-(3.4) and (3.6) hold in this case, but $s_{k,m,n}$, $r_{k,m,n}$, $\alpha_{k,m}$ and $\rho_{k,m,n}$ are replaced with $s_{k,n}$, $r_{k,n}$, $\alpha_{k,n}$ and $\rho_{k,n}$, respectively. Second, at time slot t, there are M_k available video packets of frame f for user k. Third, $\alpha_{k,n}$ is no longer pre-determined as in chapter 3 and it is one of the optimizing parameters here. Hence, we let overall coding loss factor as a matrix $\boldsymbol{\alpha}$ with $[\boldsymbol{\alpha}]_{k,n} = \alpha_{k,n}$.

To protect the information data, a channel code of fixed rate $\mathcal{B}/\mathbb{C}$ is added. Therefore, the raw video data $\tilde{r}_{k,n}[t]$ which could be transmitted is given by

$$\tilde{r}_{k,n}[t] = (\mathcal{B}/\mathbb{C}) * r_{k,n}[t] \ \forall k, \forall n. \tag{4.1}$$

Then, the modulation level on subcarrier n for queue k can be written as

$$\Gamma_{k,n}[t] = 1 + \alpha_{k,n}[t]\gamma_{k,n}[t]s_{k,n}[t] \ \forall k, \forall n. \tag{4.2}$$

(Qiu & Chawla, 1999). Note that only a finite set of modulation levels is available in practice. However, to make the optimization problem considered tractable, we first allow $\Gamma_{k,n}[t]$ to be a real number. The constraint of discrete modulation levels will be brought back in section 4.4.4.

With the CSI and the VCI that collects $R_{k,m}$, $Q_{k,m}$, $T_{k,m}$ and $\bar{\omega}$, the proposed strategy determines which packet of which user to be transmitted, assigns each subcarrier to the scheduled users, and tunes the transmission power as well as modulation level on each subcarrier once every time slot. The objective is to maximize the average quality of the received video streams without violating the playback and resource constraints.

Let us now scrutinize the BICM technique that is performed in isolation for each user queue. At the transmitter side, the selected video packets are first merged into a bitstream which will be convolutionally encoded into coded bits and then interleaved across several subcarriers. The number of interleaved bits for each subcarrier depends on the assigned M-QAM modulation level. After mapping, the BS sends the modulated symbols. At the receiver side, first, hard-decision demodulation is performed on the received signal. The demodulated bits are then de-interleaved and Viterbi decoded. Finally, the received bitstreams are unmerged into video packets which will be passed to the video decoder.

4.4 Three-Level Video Transmission Framework

This section describes the three-level strategy. First, the packet scheduling scheme is discussed. Second, an optimization problem for assigning subcarriers to the scheduled packets of different users is formulated. Finally, the bit and power allocation technique is elucidated.

4.4.1 Video Packet Importance Level $Q_{k,m}$

4.4.1.1 Optimal $Q_{k,m}$ Estimation

The importance level of each packet $Q_{k,m}$ is estimated by the amount of video quality enhancement contributed by the transmission of that packet. For each packet measurement, rather than assuming all bits of already transmitted packets are correctly received as in (Pahalawatta et al., 2007), we explicitly consider the propagation behavior of transmission errors on that packet with the past decoding information $\bar{\omega}$.

Let $\Pi_k(f, m-1) \triangleq \{\pi_{k,1}(f), \pi_{k,2}(f), \dots, \pi_{k,m-1}(f)\}$ be the reordered set of packets in queue k such that $\pi_{k,1}(f)$ is the first packet to be transmitted of current frame f for user k. Then, we define $Q_{k,m}$ as

$$Q_{k,m} = D^L_{k,f,m} - D^T_{k,f,m} \tag{4.3}$$

where $D^L_{k,f,m} \triangleq D_k\left(\tau_{k,m} = 0 \middle| \Pi_k(f, m-1), \bar{\omega}\right)$ and $D^T_{k,f,m} \triangleq D_k\left(\tau_{k,m} = 1 \middle| \Pi_k(f, m-1), \bar{\omega}\right)$ are the estimated distortion caused by the lost and the transmission of packet m after the transmission of $\Pi_k(f, m-1)$ with $\bar{\omega}$, respectively. $\tau_{k,m} \in \{0,1\}$ indicates whether packet m is transmitted $(\tau_{k,m} = 1)$ or not $(\tau_{k,m} = 0)$for user k. The distortion is measured in mean squared error (MSE) between the reconstructed and error-free frames as it is a widely used distortion metric (J. Klaue, Rathke, & Wolisz, 2003).

To clearly illustrate the concept of $Q_{k,m}$ estimation and to highlight the significance of $\bar{\omega}$, we provide an illustrative scenario in figure 4.4. In this figure, $Q_{k,m}$of a packet (white square) of the 62nd frame is estimated, with the ESD and PSD schemes. Without loss of generality, we assume that the 61st frame is the reference frame for decoding the 62nd frame and the error concealment employed by the video decoder is the TR concealment method. Then, we assume that the co-located slice of the 61st frame is somehow corrupted where the first bit error is located at the 6th MB (red square).

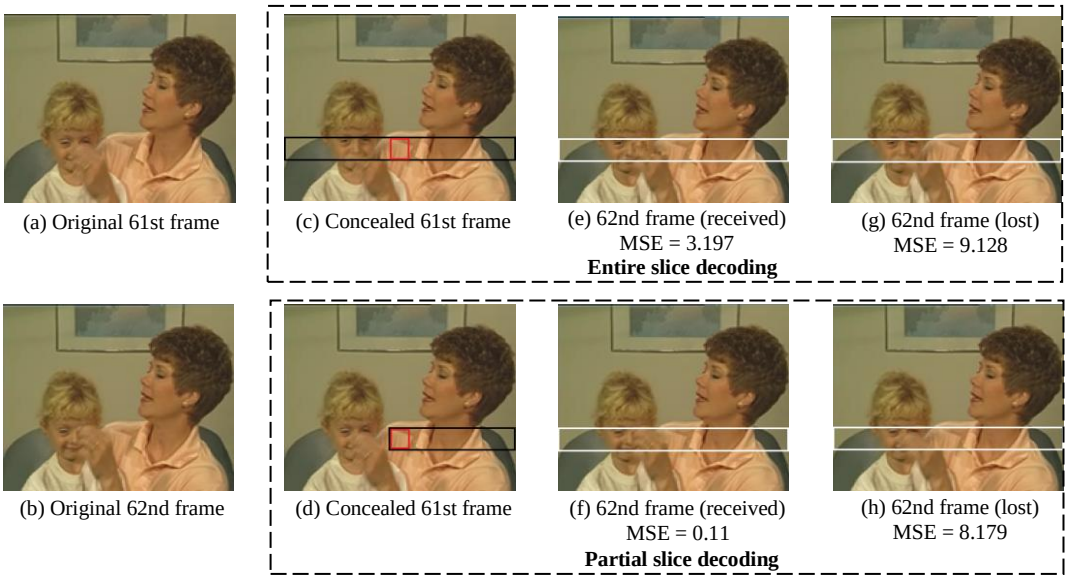


Figure 4.4 Illustrative description of the $Q_{k,m}$ estimation for a slice of the 62nd frame of *mother and daughte*r.

(a) Error free decoded 61st frame. (b) Error free decoded 62nd frame. (c) and (d): Error concealment for the 61st frame. (e) and (f): If the slice of the 62nd frame is transmitted error free. (g) and (h): If the slice of the 62nd frame is dropped.

With the ESD approach, the whole slice is discarded and concealed by co-located slice from the 60th frame (black square in (c)). In contrast, the PSD method preserves the unharmed MBs (1st to 5th MB) which can be significant to the reconstructed video quality and only conceals those affected MBs (black square in (d)). Based on the past decoding information (i.e., $\omega_{k,m}(61) = 5$ in PSD case and $\omega_{k,m}(61) = 0$ in ESD case), we estimate $Q_{k,m}$of the slice (white square) of the 62nd frame. If packet m is received correctly, the received video frame may suffer distortion due to error propagation. On the other hand, if packet m is dropped, the frame may also be distorted due to error concealment. Based on (4.3), $Q_{k,m}$ for the PSD and the ESD methods are 8.069 and 5.931, respectively. This implies that the importance of packet m from the PSD perspective is higher than that of the ESD. This is because an inter-coded packet will not be very useful to the reconstructed video quality given a distorted reference packet.

As the BS simulates the frame decoding behavior twice in order to estimate $D_{k,f,m}^{L}$and $D_{k,f,m}^{T}$ for each packet, the computational complexity is $2KM_k xy$ where x and y are the width and the height of the video frame respectively. Assuming quarter common intermediate format (QCIF), the complexity is $2KM_k(176 \times 144)$. It is true that such $Q_{k,m}$ estimation poses significant computational burden at the BS and huge signaling overhead $\bar{\omega}$, which limits the feasibility for real-time scenario. Thus, a fast yet efficient $Q_{k,m}$ estimation is desirable.

4.4.1.2 Suboptimal $Q_{k,m}$ Estimation

In this section, a suboptimal $Q_{k,m}$ estimation that eliminates the frame decoding and the $\bar{\omega}$ signaling at the BS is presented. We utilize the mathematical models as in (F. Li & Liu, 2009) to predict the MB-level transmission distortion $D_{k,f,m,mb}^{L}$ and $D_{k,f,m,mb}^{T}$,

$$D_{k,f,m,mb}^{L} = \begin{cases} W_{intra} + D_{k,f-1,m,mb}, & \text{if MB is intra} \\ \frac{1}{2} W_{inter} \sum_{j=1}^{blocks} \left(\frac{\left| MV_{mb,j}^{x} \right|}{X} + \frac{\left| MV_{mb,j}^{y} \right|}{Y} \right) + D_{k,f-1,m,mb}, & \text{if MB is inter} \end{cases} \quad (4.4)$$

$$D_{k,f,m,mb}^{T} = \begin{cases} 0, & \text{if MB is intra} \\ \frac{1}{A_m} \{ A_1 D_{k,f-1,m',mb} + A_2 D_{k,f-1,m'',mb} \}, & \text{if MB is inter} \end{cases} \quad (4.5)$$

With regard to equation (4.4), W_{intra}, W_{inter}, $MV_{mb,j}^{x}$,and $MV_{mb,j}^{y}$ denote the weight factor of intra MB, weight factor of inter MB, MV in x and y axes, respectively. Generally speaking, W_{intra} is larger than W_{inter} as intra MBs intuitively induce more distortion. If a MB is inter, it is further divided into 16 blocks of 4 x 4 size. Besides, X and Y represent the partition mode in which the H.264 standard supports seven MB partition modes as described in section 4.2.1. With respect to equation (4.5), A_m, A_1,

and A_2 denote the area of packet m, the areas of overlapped regions between the current MB and the reference MBs in packet m' and m'' of the previous frame (to account for error propagation), respectively. More details of this model can be found in (F. Li & Liu, 2009).

As MB-level transmission distortion is additive (Y. F. Zhang et al., 2010), $D^L_{k,f,m}$ and $D^T_{k,f,m}$ in equation (4.3) can be predicted from $D^L_{k,f,m} = \sum_{mb=1}^{MB_{k,f,m}} D^L_{k,f,m,mb}$ and $D^T_{k,f,m} = \sum_{mb=1}^{MB_{k,f,m}} D^T_{k,f,m,mb}$, respectively. $MB_{k,f,m}$ is the number of MBs in packet m of frame f of user k. Obviously, all parameters from equations (4.4)-(4.5) can be readily obtained during the video encoding process except the distortion of the previous frame $f-1$ that can only be computed on the fly. This enables us to eradicate the exhaustive video decoding process.

Consider first the PSD case, if packet m is transmitted, the expected distortion of packet m is given by

$$D_{k,f,m} = \left\{D^L_{k,f,m} P_{k,f,m,1}\right\} + \left\{\sum_{mb=1}^{MB_{k,f,m}-1}\left(\sum_{g=1}^{mb} D^T_{k,f,m,g} + \sum_{h=mb+1}^{MB_{k,f,m}} D^L_{k,f,m,h}\right) P_{k,f,m,mb+1} \times \prod_{i=1}^{mb}\left(1-P_{k,f,m,i}\right)\right\} + \left\{D^T_{k,f,m} \prod_{i=1}^{MB_{k,f,m}}\left(1-P_{k,f,m,i}\right)\right\} \quad \forall k, \forall m \tag{4.6}$$

where $P_{k,f,m,mb} \triangleq 1-\left(1-BER_{k,f,m,mb}\right)^{R_{k,f,m,mb}}$ is the probability that a bit error occurs in MB mb and $R_{k,f,m,mb}$ is the length of MB mb (in bits). The first term corresponds to the expected distortion when the first bit error occurs at MB 1. The second term is the case when the first bit error occurs at MB $mb+1$. Thus, all MBs

following MB $mb+1$ are lost and concealed as discussed earlier. The last term refers to the case when there is no bit error.

Now, with $\left(D_{k,f,m}^{L}\right)$ replacing $\left(\sum_{g=1}^{mb} D_{k,f,m,g}^{T} + \sum_{h=mb+1}^{MB_{k,f,m}} D_{k,f,m,h}^{L}\right)$ in the second term of equation (4.6), this gives rise to the ESD case. It implies that any bit error would render the entire packet being discarded and concealed. With equation (4.6), we can compute $Q_{k,m}$ in equation (4.3), which requires $\mathcal{O}\left(KM_k\left(MB_{k,f,m}+1\right)\right)$ operations. Indeed, the continuous update of the information associated with previous frame $f-1$ captures the decoding error that accumulates over time. Thus, it eliminates the $\bar{\boldsymbol{\omega}}$ signaling. Note that if packet m is not transmitted before exceeding its deadline, $D_{k,f,m}$ is set to $D_{k,f,m}^{L}$. Note also that distinguishable from (F. Li & Liu, 2009), our work here explicitly considers the effect of bit errors on each MB and the BER value is dynamic (depending on our optimization) rather than predetermined. The steps on how to obtain $BER_{k,f,m,mb}$ and its corresponding $P_{k,f,m,mb}$ will be revealed in section 4.4.4.

4.4.2 Packet Scheduling

The goal of packet scheduling is to sort the video packet transmission order across all users. Unlike existing scheduling works that usually ignore the transmission reliability of video packets, an *overall target BER* is considered here to control how much average rate (packets) can be transmitted during each time slot. The rate is obtained by substituting $BER_{k,n} = \textit{overall target BER}$ into equations (3.4) and (3.3). From the BS perspective, it would seem logical that the packets are forwarded as fast as possible in order to avoid deadline violations. However, the pitfall is that many received packets are possibly corrupted and thus undecodable. On the other hand, if packets are forwarded in a conservative way, it may render limited achievable video quality though

most of the packets are useful. This is in line with the observation in (D. W. Wang et al., 2013), that *overall target BER* has a critical impact on the video performance and its optimal value cannot be known a priori. In this regard, resource allocation should be adaptive to the scheduled packet priority in an UEP way so that the average overall video distortion can be minimized.

4.4.3 Subcarrier Assignment

To bridge the gap, we propose a subcarrier assignment strategy that assigns stricter target BERs to more important packets and *vice versa*. To this aim, we formulate an optimization problem that minimizes a weighted sum of user coding gaps subject to the scheduled packet length constraints. Assuming EPA (i.e., $s_{k,n} = P_{tot}/N$), the problem can be expressed as

$$\min_{\{\boldsymbol{\tau},\boldsymbol{\rho},\mathbf{s},\boldsymbol{\alpha}\}} \sum_{k=1}^{K}\sum_{n=1}^{N}\left(W_k[t]\rho_{k,n}[t]\alpha_{k,n}[t]\right)$$

$$\text{s.t.} \quad (3.2),(3.3),(3.6), \quad \sum_{n=1}^{N}\rho_{k,n}\, r_{k,n}[t] \geq \frac{\mathbb{C}}{\mathcal{B}} R(\boldsymbol{\tau}_k) \quad k = 1,2,\dots,K \tag{4.7}$$

where W_k denotes the weight for user k, $\boldsymbol{\tau} = \{\boldsymbol{\tau}_1, \boldsymbol{\tau}_2, \dots, \boldsymbol{\tau}_{\mathrm{K}}\}$ is the set of transmission policy for all users and $\boldsymbol{\tau}_{\mathrm{k}} = \left(\tau_{k,1}, \tau_{k,2,}, \dots, \tau_{k,M_k}\right)$ is the vector of transmission policy for all packets for user k. Recall that $\tau_{k,m}$ is directly obtained from the output of the packet scheduling phase. The fourth constraint implies that the radio resources allocated to user k should be adequate enough to transmit the video packets under $\boldsymbol{\tau}_k$, where $R(\boldsymbol{\tau}_k) = \sum_{m=1}^{M_k} \tau_{k,m} R_{k,m}$. Hereafter we omit the time index t for simplicity as it remains invariant throughout this section. We then project the importance level of video packets into the weight, which is given by

$$W_k = \sum_{m=1}^{M_k} \tau_{k,m} Q_{k,m} \qquad \forall k \tag{4.8}$$

It is apparent that a group of more important packets from user k will lead to a larger W_k. Thus, user k has higher priority to occupy a subcarrier such that their corresponding $\alpha_{k,n}$ is reduced. However, user k does not essentially dominate all the best subcarriers. This may happen when another user who happens to have only a few good subcarriers that overlap with those of user k. Furthermore, the link reliability of a packet is tightly connected with the modulation order adopted to carry that packet. If user k already obtains sufficient subcarriers (with low modulation orders) to reliably transmit all data, remaining subcarriers can be granted to other users.

All these essences are captured in problem (4.7) and solving it actually translates into a reduction in the average distortion of all scheduled packets. As $\alpha_{k,n}$ is an increasing function of $BER_{k,n}$, solution to (4.7) yields to situation where more important packets are possibly error-free due to stricter (smaller) BERs on the one hand, and less critical packets are exposed to higher BERs on the other hand. Fortunately, the amount of quality degradation for those low-priority packets is anticipatively minimal as they are easily concealable (if they are distorted). The relationship between $\alpha_{k,n}$ and $BER_{k,n}$ is shown in appendix A.

Note that problem (4.7) is solved in every time slot and so does the update of W_k associated with $Q_{k,m}$. Since $Q_{k,m}$ depends on the past scheduling decisions, solution to (4.7) offers means of effecting long-term fairness among different users in the following sense. If scheduled packets of user k are assigned loose target BER in the current time slot, the current frame may suffer distortion that would propagate to subsequent frames due to the use of predictive coding. In response to the loss, $Q_{k,m}$ in new time slot would

essentially grow larger. This enforces the BS to assign stricter BER to user k's packets in the next time slot.

Denote μ_k as the Lagrange multipliers for the rate constraints and let $[\boldsymbol{\mu}]_k = \mu_k$ be the multiplier vector. Then, the optimization problem in (4.7) is equivalent to finding the minimum of the following Lagrange function

$$\begin{aligned}\mathcal{L}(\boldsymbol{\rho}, \boldsymbol{\alpha}, \boldsymbol{\mu}) = &\sum_{k=1}^{K}\sum_{m=1}^{N}\left(W_k \rho_{k,n}\alpha_{k,n}\right) \\ &-\sum_{k=1}^{K}\mu_k\left(\sum_{n=1}^{N}\rho_{k,n}r_{k,n} - \left(\frac{\mathbb{C}}{\mathcal{B}}\sum_{\mathrm{m}=1}^{M_k}\tau_{k,m}R_{k,m}\right)\right)\end{aligned} \tag{4.9}$$

Comparing equation (4.9) with (3.12), we can observe that these two equations share the same basic form and therefore problem (4.7) can be solved in a similar fashion as in section 3.3.1. Utilizing those derivations allow us to reach the point where

$$\alpha'_{k,n} = \left[\frac{\tau\mu_k}{T_{ofdm}\ln 2} - \frac{1}{\gamma_{k,n}s_{k,n}}\right]^{+} \forall k, \forall n, \tag{4.10}$$

and

$$\rho'_{k,n} = \begin{cases}1, & k = \arg\min\limits_{k} F_{k,n}\left(\alpha'_{k,n}\right) \\ 0, & \text{otherwise.}\end{cases} \quad \forall n. \tag{4.11}$$

. Then by replacing the terms "(3.16)" and "(3.17)" with "(4.10)" and "(4.11)", respectively, algorithm 3.1 can again be executed to obtain the solution of problem (4.7).

4.4.4 Bit and Power Allocation

As explained earlier, we first loosen the constraint of discrete values on the modulation levels to obtain a closed-form solution. For discrete modulation level, we pick the constellation size for each subcarrier in a greedy way (G. C. Song & Y. G. Li, 2005).

Since we adopt square MQAM constellations, we allow $\Gamma_{k,n}$ to take value from the following set

$$\Gamma_{k,n} \in \{0, \dots, 2^b, \dots, 2^{b_{max}}\} \, \forall k, \forall n. \tag{4.12}$$

where b represents an even number starting from $b = 2$ and stands for the number of bits. b_{max} is the maximum number of bits and its corresponding $2^{b_{max}}$ is the highest modulation level. Note that $\Gamma_{k,n} = 0$ indicates no transmission.

Based on the objective function in (4.7) with the assumption of EPA, an obvious way is to assign next modulation level to the subcarrier which requires the lowest additional coding loss factor $\alpha_{k,n}$. However, it may not yield good video quality as the Viterbi decoder sees the whole subchannel (those subcarriers that have been assigned to user k as a channel with fluctuating BERs associated with $\text{SNR}_{k,n} = s_{k,n} \times \gamma_{k,n}$. It is well known that channel coding performs poorly in such environments and frequency-selective power allocation (FPA) can enhance the performance (Moon & Cox, 2009).

Inspired by this fact, FPA is also employed here, but with an entirely different means. The focus here is on multiuser setting and adaptive modulation rather than single-user scenario and static modulation as in (Moon & Cox, 2009). The goal is to generate constant SNR across those subcarriers with the same modulation order such that the potential channel coding gain is harnessed. This is accomplished by adopting the adaptive bit loading algorithm in (G. C. Song & Y. G. Li, 2005). The difference is that for each user, the algorithm simply runs until the target data rate is reached and the total power is normalized to the total power that has been assigned in the subcarrier assignment phase. Recall in that phase, more subcarriers are assigned to more important packets and so does the assigned total power for that user. Hence, the nature of protecting high-priority packets preserves. The whole steps are given in algorithm 4.1.

Table 4.2 Algorithm 4.1 Bit and power allocation.

1: Initialization: set $e_k, b_{k,n}$, and $s_{k,n} = 0\ \forall k, \forall n$, where e_k denotes the number of bits needed to transmit the selected packets from queue k. Compute $\alpha_{k,n}$ according to (3.4) with $BER_{k,n} = overall\ target\ BER$ that has been set in packet scheduling stage.

2: Define $f(b_{k,n}, \gamma_{k,n}, \alpha_{k,n})$ as $(2^{b_{k,n}} - 1)/(\gamma_{k,n}\alpha_{k,n})$ according to (4.2).

3: **for** k=1 to K

4: set $S_k = 0$

5: **repeat**

6: **for** all subcarrier $n \in \psi_k$

7: **If** $b_{k,n} < b_{max}$ (the maximum modulation level is not reached)

8: Compute $\Delta s_{k,n} = f(b_{k,n} + 2, \gamma_{k,n}, \alpha_{k,n}) - f(b_{k,n}, \gamma_{k,n}, \alpha_{k,n})$

9: **else**

10: $\Delta s_{k,n} = \infty$.

11: **end for**

12: $n' \leftarrow \arg\min_n \Delta s_{k,n}$

13: set $s_{k,n'} = s_{k,n'} + \Delta s_{k,n'}$, S = S + $\Delta s_{k,n'}$, $b_{k,n'} = b_{k,n'} + 2$,and $e_k = e_k + 2$

14: **until** $e_k = (\mathbb{C}/\mathcal{B}) \sum_{m=1}^{M_k} \tau_{k,m} R_{k,m}$

15:: $s_{k.n} = s_{k.n} * (P_{tot}/N) * |\psi_k|/S_k$ such that S_k is normalized to $(P_{tot}/N) * |\psi_k|$ $\forall n \in \psi_k$

16: **end for**

In line 12, we need to determine the subcarrier with the smallest power increment, which requires $|\psi_k|$ comparisons in each of the $e_k/2$ iterations. Therefore, the total computational complexity is $\mathcal{O}((e_1|\psi_1| + e_2|\psi_2| + \cdots + e_k|\psi_K|)/2)$.

Recall that if suboptimal $Q_{k,m}$ estimation is used, we need to acquire $BER_{k,f,m,mb}$, which is quite straightforward. As we already have $\Gamma_{k,n}$, $\gamma_{k,n}$, and $s_{k,n}$ by executing algorithm 4.1, equation (4.2) allows us to obtain $BER_{k,n}$. We then substitute $BER_{k,f,m,mb} = BER_{k,n}$ for all MBs of user k.

The power allocation strategy deserves further attention. So far in the OFDM resource allocation literature, EPA is the most commonly used assumption to alleviate the computational complexity. While such simplification enjoys the (near) optimal performance in efficiency-oriented resource allocation policies, this is not the case for fairness-oriented schemes (H. W. Lee & Chong, 2008). The reason is that when fairness is emphasized, some bad channel users have to be scheduled that culminates in high channel gain variations among the user. In this case, FPA is necessary as the achieved performance gain can be significant (Bohge, Gross, & Wolisz, 2005). However, these works justify their own arguments from the theoretical perspective and isolate the influence of channel coding. In contrast, this chapter investigates the necessity of FPA from the video performance perspective.

4.5 Simulation Results

In the simulations, six video sequences with 100 frames: *foreman*, *carphone*, *coastguard*, *silent*, *mobile* and *news* in QCIF format were used. The video sequences were encoded using JVT reference software 18.4 ("H.264 JM 18.4 Reference Software,"), at the baseline profile, 8 bpp, a frame rate of 25 fps and an encoding rate of 350 kbps. All frames, except for the first frame, were encoded as P frames. To enhance error robustness, random I MBs were inserted into each frame, and **Constrained Intra Prediction** was activated. The playback delay T_{max} of each packet was set to 40 ms. The TR concealment method were adopted at the H.264 decoder which had been modified to perform PSD.

For the OFDMA system, the total bandwidth B_T = 1 MHz, the total transmit power P_{tot} = 1 W, the number of subcarriers N = 256 and the time slot duration τ = 5ms corresponding to 19 OFDM symbols. We assume that 15 out of 19 OFDM symbols were adopted for packet transmission while the rest was supposed to account for the signaling overhead. The channel coding parameters are the same as in (D. W. Wang et al., 2013), in which the convolutional encoder employs a rate of $\mathcal{B}/\mathbb{C} = 1/2$ with constraint length $\Theta = 5$ and code generator polynomial of [23,35] in octal. Hard-decision Virterbi decoding was performed at the receiver. We assume four different MQAM constellations with Gray bit mapping, that is $\Gamma \in \{0, 4, 16, 64, 256\}$ corresponding to 0, 2, 4, 6, and 8 bits per symbol. For each subcarrier n, individual subcarrier $SNR_{k,n} = s_{k,n} \times \gamma_{k,n}$ was computed and channel errors were inserted into the modulated bitstream according to $SNR_{k,n}$. The average SNR of all users was set to 15 dB. The rest of the wireless system parameters are the same as in section 3.4.2.

For suboptimal $Q_{k,m}$ estimation, the parameters are the same as in (F. Li & Liu, 2009) except that we empirically set W_{inter} to 0.1. For the packet scheduling (upper level of the framework), we adopted the state-of-the-art packet scheduling scheme as proposed in (F. Li et al., 2012). We refer this scheme to joint-allocation (JA) scheme as it simultaneously schedules packets and assigns subcarriers to the selected packets. Benefiting from the joint allocation, the JA scheme has significantly outperformed many existing scheduling schemes (including (Ji et al., 2009; Pahalawatta, 2007)), as reported in (F. Li et al., 2012). This is one reason why we select the JA. Note that we do not intend to develop a new packet scheduling scheme, but for the sake of completeness, we briefly describe the JA process as follows. The packet with the best ratio of $Q_{k,m}/J_{k,m}$ is selected in one loop, where $J_{k,m}$ denotes the number of remaining unassigned subcarriers needed to transmit packet m of user k. This loop repeats until no more free subcarriers can carry any packet. During each loop, computing $J_{k,m}$ requires

two steps. First, for each user, we sort the remaining subcarriers in descending order of $\gamma_{k,n}$. Second, with equal power distribution over all subcarriers, the corresponding transmission rate of each subcarrier is computed with equation (3.3) and then added along the sorted subcarrier sequence until the aggregate rate is large enough to transmit the packet. As explained earlier, the JA now has to consider the *overall target BER* in the transmission rate computation in which it neglects. Note that switching the *overall target BER* may change not only the number of scheduled packets, but also the packet order to be transmitted. This is because $J_{k,m}$ is associated with the transmission rate that varies accordingly with the *overall target BER*. Overall, the computational complexity of this stage is $N(K\log_2 N + KM_k)$.

To assess the decoded video quality, we use PSNR $\triangleq 10\log_{10}\left(\left(2^{\text{bpp}}-1\right)^2/\text{MSE}\right)$ as it is the most commonly used metric to score the quality of encoded video (J. Klaue, Rathke, & Wolisz, 2003). For each video, the PSNR as reported below is averaged over 100 frames and over five sets of decoded video corresponding to five random error patterns for each time slot. To better assess the effectiveness of different components (i.e., decoder-based error resilience, packet importance estimation, subcarrier assignment and power allocation) in the proposed framework, we compare six algorithms as follows.

1. BER-JA: First, $Q_{k,m}$ is calculated based on optimal$Q_{k,m}$ estimation with $\bar{\omega}$ feedback. Second, packet scheduling is performed based on JA. Then, algorithms 1 and 2 execute such that subcarriers are allocated to each selected users' packets according to $Q_{k,m}$ and frequency-selective power is distributed over all the assigned subcarriers. Finally, the video decoder performs PSD.

2. BER-JA-EP: This is the same as BER-JA except that EPA is distributed.

3. BER-JA-PREDICT: This is the same as BER-JA except that $Q_{k,m}$ is calculated based on suboptimal $Q_{k,m}$ estimation.

4. BER-JA-WF: This is the same as BER-JA except that $Q_{k,m}$ is calculated based on optimal $Q_{k,m}$ estimation without $\bar{\omega}$ feedback. This is achieved by assuming all transmitted MBs are perfectly received (i.e., set all $\omega_{k,m}$= number of MBs in a slice). Note that we design this algorithm is to ensure fair comparisons with the following scheme.

5. JA-WF (F. Li et al., 2012): The first two steps are the same as those of BER-JA-WF. However, algorithm 1 does not execute as it directly assigns subcarriers to each user during the packet scheduling. Hence, only one target BER is considered. Finally, equal-power allocation is employed.

6. MAX (Jang & Lee, 2003): This scheme serves as a benchmark. It is a content-blind approach that maximizes the sum rates of all users. This is achieved by assigning each subcarrier to the user with the maximum channel gain on it. It leads to a complexity of $\mathcal{O}(KN)$.

The main differences among each scheme are summarized in table 4.3.

For clarity of exposition, we refer to the proposed schemes with the prefix "BER" as BER-aware schemes.

Figure 4.5 shows the average PSNR over all six video sequences for different *overall target BER*s and methods. We partition the discussion into four parts to facilitate the analysis.

Table 4.3 Properties for different algorithms.

Scheme	$Q_{k,m}$ Estimation	Error Resilience	$\bar{\boldsymbol{\omega}}$	Computational Complexity at BS
BER-JA	Optimal	PSD	Yes	$C_1+C_2+C_4+C_5+\bar{\boldsymbol{\omega}}$ delay
BER-JA-EP	Optimal	PSD	Yes	$C_1+C_2+C_4+C_5+\bar{\boldsymbol{\omega}}$ delay
BER-JA-PREDICT	Suboptimal	PSD	No	$C_3+C_2+C_4+C_5$
BER-JA-WF	Optimal	ESD	No	$C_1+C_2+C_4+C_5$
JA-WF	Optimal	ESD	No	$C_1+C_2+C_5$
MAX	No	ESD	No	C_6+C_5

$C_1 = \mathcal{O}(2KM_k xy)$; $C_2 = \mathcal{O}\big(N(K\log_2 N + KM_k)\big)$; $C_3 = \mathcal{O}\left(KM_k\big(MB_{k,f,m}+1\big)\right)$;

$C_4 = \mathcal{O}(Z_1 KN)$; $C_5 = \mathcal{O}((e_1|\Omega_1| + e_2|\Omega_2| + \cdots + e_k|\Omega_K|)/2)$; $C_6 = \mathcal{O}(KN)$

1. Comparisons under Different Overall Target BERs

It can be seen that the PSNR curves for all schemes possess different optimal *overall target BERs*. Nevertheless, when the *overall target BER* is stringent (10^{-3}), the performances of almost all schemes tend to drop. This is because the number of video packets available at the video decoder is fewer, thereby limiting its achievable video quality. On the other hand, when the *overall target BER* is loose (10^{-1}), more packets fill the video buffer but mostly are not decodable due to the presence of huge amount of bit errors. It is evident that BER-aware schemes outperform their counterparts for the entire BER range under the same settings of error resilience and $Q_{k,m}$ estimation. This implies that without knowing the optimal *BER* to operate with, the proposed scheme will always be a safer option due to its robustness against channel errors.

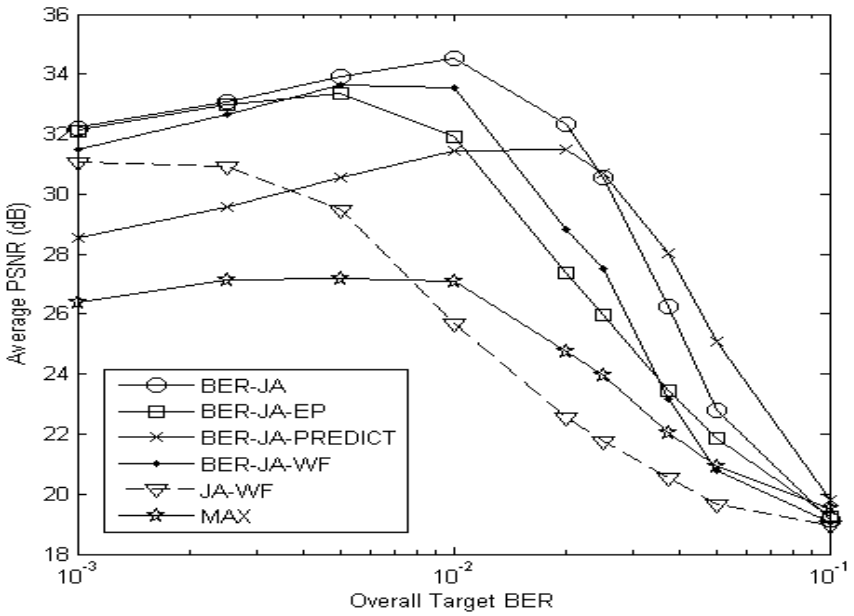


Figure 4.5 Average PSNR over all users versus *overall target BER* for different schemes.

In addition, it is worth noting that the performances reached by JA-WF and MAX are different from those as reported in (F. Li et al., 2012), i.e., JA-WF does not always outperform MAX. The reason is that we have analyzed the video performance in a more complete and complex environment (bit-level simulation) with respect to those simulated in (F. Li et al., 2012). At high *overall target BER* region ($1\sim10\times10^{-2}$), the performance of JA-WF is worse than that of MAX. This is because the burst error stemming from equal power allocation severely degrades the performance of channel coding. On the other hand, at low *overall target BER* region ($1\sim10\times10^{-3}$), less bit errors occurs and so does the degree of burst error pattern. At this region, JA-WF that prioritizes the transmission of important video packets could gain the upper hand. In brief, MAX exhibits the worst overall performance as it solely pursues the highest network capacity by ignoring the video content. It is 1.6-7.6 dB worse than BER-JA.

2. Comparisons with JA-WF in (F. Li et al., 2012)

Prior to discussion, we remark that BER-JA-WF and JA-WF are transmitting the same video packets but with different subcarrier assignment and power allocation strategies.

Thus, the performance gain is purely derived from the lower level of the proposed framework. From figure 4.6, it can be observed that BER-JA-WF remarkably outperforms JA-WF, reaching up to 7.9 dB. This is contributed by our scheduling prioritization strategy that assigns different target BERs to different video packets whereas JA-WF treats all packets equally.

3. Optimal $Q_{k,m}$ Estimation vs. Suboptimal $Q_{k,m}$ Estimation

As expected, BER-JA with optimal $Q_{k,m}$ estimation yields overall better performance than BER-JA-PREDICT with suboptimal $Q_{k,m}$ estimation. The PSNR gain up to 3.7 dB is derived from the accurate prediction of packet importance level. However, it can be observed that the gain becomes negative at BER region of $(3.75{\sim}10 \times 10^{-2})$. The reason for this ramification lies as follows. At this region, transmission rate during each time slot $r_{k,n}[t]$ is much higher, thus enabling many video packets to be forwarded at the same slot. With optimal $Q_{k,m}$ estimation, the video packets scheduled at earlier time slots are really those of utmost importance due to accurate prediction. This renders very few less important packets left in the transmission queue during latter time slots. In this case, radio resources are adequate to protect these packets but the quality contributions made by these packets are restricted. On the contrary, suboptimal $Q_{k,m}$ estimation seems benefiting from the mismatch as those remaining packets during the latter time slots may have higher actual $Q_{k,m}$.

4. Frequency-Selective vs. Equal Power Allocation

It is apparent that frequency-selective power allocation provides substantial quality improvement over equal power allocation, reaching up to 4.9 dB. Thus, in the context of practical communication systems where channel coding is actively utilized, equal power allocation could be an inappropriate choice and frequency-selective power allocation is necessary to fully exploit the channel coding performance.

Figure 4.6 depicts the average PSNR for each video corresponding to *overall target BER* = 5×10^{-3}. It is obvious that most of the BER-aware schemes are superior to their counterparts over all sequences. A larger performance improvement can be consistently accrued in high-motion sequences such as *foreman* and *carphone*. This is because the packets with complex content are generally more important that that with simple content. Thus, these packets are likely to be scheduled where the benefit of protecting these packets via stricter target BERs can be more visible. On the other hand, when the packets from slow-motion sequences such as *news* are scheduled, they are extremely important since many inter-coded packets rely on them. The effect of error propagation can last for a long period if these packets are distorted. BER-aware schemes recognize this fact by assigning stricter targets to them.

Figure 4.7 depicts the average PSNR and the variance of the PSNR per frame over all users for different schemes corresponding to *overall target BER* = 1×10^{-2} . Due to lengthy restriction, we only provide the results from the 10^{th} to 90^{th} frame. From figure 4.7 (a), it is obvious that BER-aware schemes perform better than their counterparts.

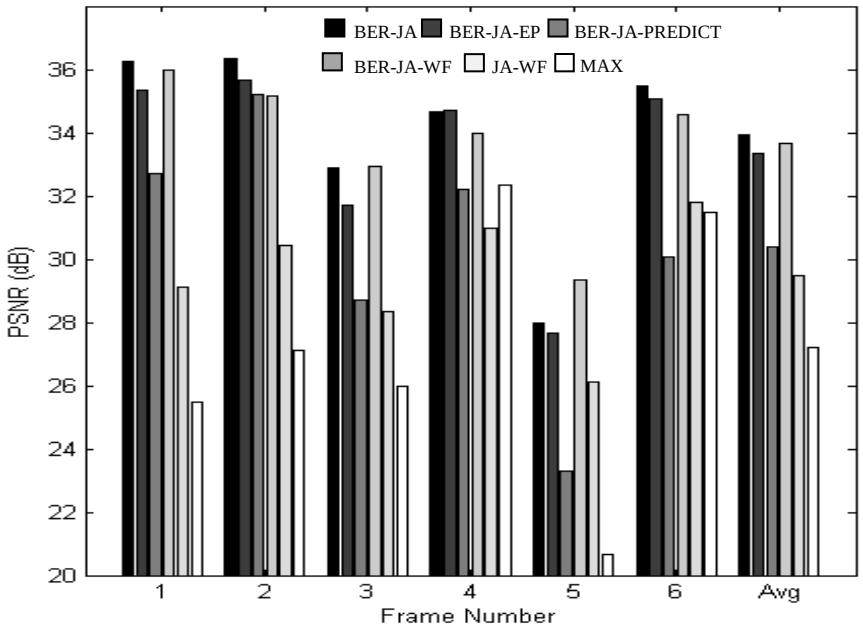
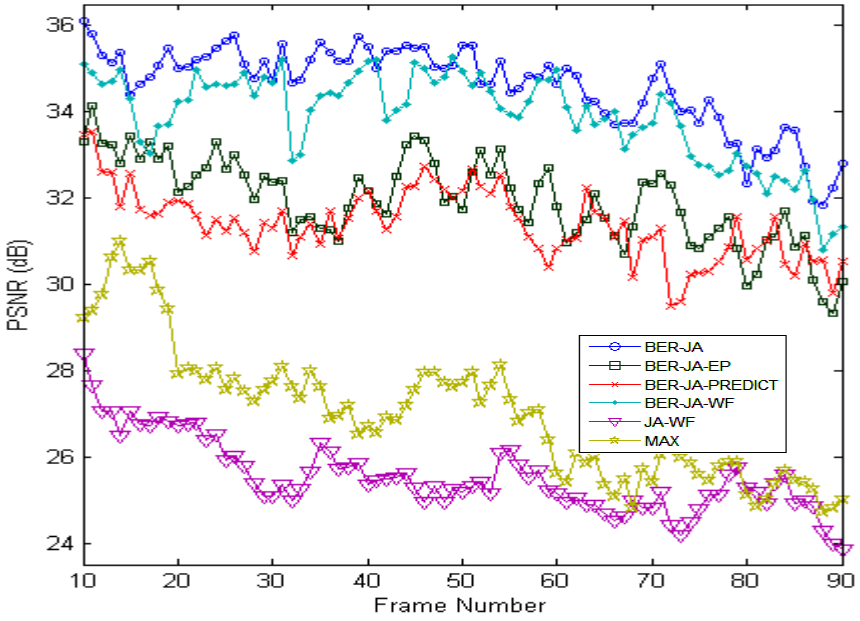
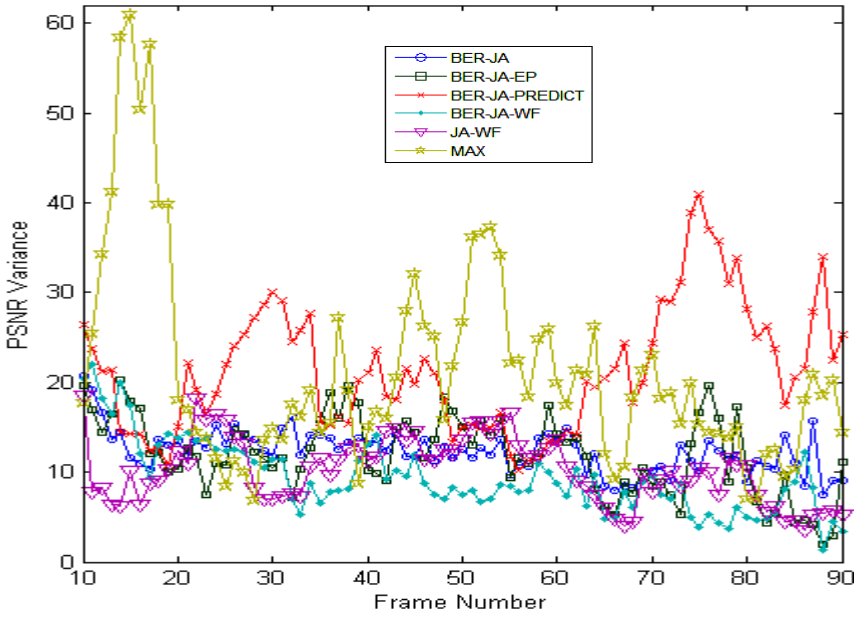


Figure 4.6 Average PSNR per video for different schemes at *overall target BER* = 5×10^{-3}. Sequence number represents 1: *foreman*; 2: *carphone*; 3: *coastguard*; 4: *silent*; 5: *mobile* and 6: *news.*

From figure 4.7 (b), we can observe that BER-JA, BER-JA-EP, BER-JA-WF and JA-WF tend to provide similar video quality per frame across all the users. The variances of these four schemes are quite close to each other. This is a very meaningful and important observation as it implies that our proposed schemes do not penalize the performance of certain user in order to maximize the total video quality. In addition,



(a)



(b)

Figure 4.7 Frame-by-frame quality over all users for different schemes at *overall target BER* =$\mathbf{1 \times 10^{-2}}$.
(a) Average PSNR per frame. (b) PSNR variance across all users.

BER-JA-PREDICT is better than WF, but worse than JA-WF in terms of video-quality fairness. Such performance degradation is acceptable considering that the complexity is remarkably reduced according to table 4.2. Moreover, the relatively larger PSNR variance is attributed to the fact that BER-JA-PREDICT yields at least 4.6 dB better than JA-WF in terms of frame-by-frame quality. This indicates that the proposed low-complexity algorithm is still very promising and can be used widely in real-time applications.

4.6 Summary

In this chapter, we have proposed a BER-driven resource allocation scheme for use in a nonscalable H.264 framework. We incorporated the state-of-the-art packet scheduling strategy into the upper level of the framework. At the lower level, a novel subcarrier allocation problem was formulated to provide UEP in terms of target BERs to different scheduled packets according to the packet-level importance. Bit-level simulation results show that the proposed schemes remarkably outperform the conventional schemes, with the PSNR gain as high as 7.9 dB. It is also demonstrated that a suboptimal estimation of packet importance, which reduces the complexity from $\mathcal{O}(2KM_k xy)$ to $\mathcal{O}\left(KM_k\left(MB_{k,f,m}+1\right)\right)$, is still better than the conventional schemes. We further tested its performance through the comparison with equal power allocation. The significant performance loss of using equal power allocation made us to conclude that frequency-selective power allocation is necessary to fully exploit the potential channel coding gain.

CHAPTER 5

BER-Driven Resource Allocation (BRA) for Realistic Scalable Multimedia Bitstreams

In this chapter, we discuss the application of BRA in a scalable multimedia bitstream framework. Together with the changes in the coding mode, consideration of the constraints pertaining to this case entails changes in the optimized multimedia delivery over downlink OFDMA systems. This chapter ends with a simulation study detailing the performance of the proposed scheme in terms of received quality.

5.1 Introduction

Currently, nonscalable coding is still dominating the market of multimedia streaming services (Khalek et al., 2012). Scalable coding, on the other hand, is gaining popularity at a rapid pace due to its flexible bitstream adaptation to a wide range of network conditions and terminal capabilities. Such coding, however, introduces an inherent packet ordering that restricts the possible scheduling strategies at the scheduler. That is, every successive enhancement layer contributes a certain amount of temporal/spatial/fidelity refinement only when all preceding layers are reliably received. For example, if the base layer is neither scheduled nor well received, the quality improvements contributed by the successive layers are futile. In this regard, BRA scheme can again be utilized to assign stricter target BERs to more important layers.

This chapter considers the problem of JPEG 2000 (J2K) scalable bitstream transmission over downlink OFDMA networks. Unlike the case in previous chapter, multimedia packets cannot be transmitted out of order and still can be decoded due to hierarchical decoding dependencies as stated above. Moreover, the amount of radio resources allocated to the base layer directly affects the quality of all the enhancement layers, the amount of resources given to the first enhancement layer directly affect its successive

layers, and so forth. Thus, the scheduler must adopt expected quality rather than deterministic quality during the optimization. It follows that the ultimate objective here is to maximize the expected quality of all J2K scalable bitstreams subject to a total power constraint. Since the proposed BRA techniques to date can be extended to scalable bitstreams, the chief concern in this chapter boils down to the question on how many layers (starting from the base layer) of each user should be transmitted under some channel conditions. Ideally, when channel conditions are good, more layers are sent. As connection quality degrades, we propose a *water-filling* algorithm that transmits a subset of all layers to maintain the overall transmit reliability.

Extensive research has focused on the resource allocation for scalable-bitstream transmission. In (Maani & Katsaggelos, 2010), an optimized bitstream extraction and content-aware channel rate allocation is proposed for the transmission of scalable video coded (SVC) packets. However, the system is primarily designed for a generic channel model, which is not directly applicable in an OFDM system. In the context of OFDM systems, much active research has been devoted to the technique known as joint source/channel coding (JSCC) (Chan, Cosman, & Milstein, 2006; Y. Sun & Xiong, 2006). The basic idea behind JSCC technique is to distribute source and channel coding bits among total transmit bits in such a way that more important data suffer less distortion at the expense of more distortion for less critical data.

This chapter distinguishes itself by looking at the problem from the OFDM optimization in terms of subcarrier assignment, power allocation and adaptive modulation. Some related works are the following. In (Park, Lee, Lee, & Bovik, 2009), a cross-layer approach that jointly tunes SVC source and wireless channel parameters is proposed. The authors in (Khalek et al., 2012) proposes another cross-layer design for scalable video over wireless networks. An additional parameter called playback buffer state is taken into account. The above two methods, however, consider only a single-user

scenario. While the authors in (Yang & Wang, 2013) do address the relevant problem in the context of multiuser scenario, EPA is adopted. This expectably yields poor received multimedia quality as reported in chapter 4.

In (Ahmad, Worrall, & Kondoz, 2008), an unequal power allocation (UPA) scheme for the transmission of SVC packets is proposed. In particular, more power is allocated to the base layer packets than the enhancement layer packets based on a lookup table. However, the lookup table is constructed based on a fixed modulation level in which the amount of power allocated to a layer is inversely proportional to the BER of the layer. This relationship does not hold when adaptive modulation is adopted. In (Sabir et al., 2010), the authors advocate similar strategy to enhance the performance of JPEG transmission over MIMO systems. Again, static modulation scheme is assumed.

In (S. S. Tan et al., 2008), an adaptive modulation scheme for transmitting scalable images with multiple description coding via OFDM networks is proposed. However, the total transmit rate maximization is performed without exploiting multiple target BERs. In (Su, Han, Wu, & Liu, 2006), a framework to transmit multiple scalable video bitstream over OFDM networks is proposed. Although multiple target BERs are considered, the target BER is only chosen from either 10^{-5} or 10^{-6}, depending on which BER setting achieving the lowest expected distortion. Moreover, the BER disparity only exists between the least important layer and the rest layers.

The rest of the chapter is organized as follows. Section 5.2 provides an overview of J2K coding. Section 5.3 describes our system model. In section 5.4, we propose a three-level framework consisting of dynamic target BER assignment, unequal subcarrier assignment and bit and power allocation. Section 5.5 provides simulation details and results. We wrap up the chapter in section 5.6. A list of the key mathematical notations used in this chapter is provided in table 5.1.

Table 5.1 List of Key Notations

Notation	Description
O	DWT resolution level
$E[Q]$	Expected quality
$s_{k,m,n}$	Power allocated on subcarrier n for user k
$\rho_{k,m,n}$	Indicator for subcarrier assignment
P_{tot}	Total available power
M_k	Number of layers available at the BS
M'_k	Number of queues to be transmitted
C	Number of CBs
$W_{k,c}^{max}$	Number of coding passes for CB c of user k
$\Delta r_{k,c,w}$	Rate of coding pass w in CB c of user k
$\Delta d_{k,c,w}$	Distortion reduction in terms of normalized MSE for coding pass w in CB c
$R_{k,m}$	Length of layer m of user k
$W_{k,c}^{m}$	Number of coding passes of CB c in the layers 1 to m
$LER_{k,m}$	Layer error rate for layer m of user k
$\Delta Q_{k,m}$	Quality contribution of layer m of user k
$\gamma_{k,n}$	Channel-to-noise ratio on subcarrier n for user k
$E[ch_k]$	Average subcarrier CNR
$\psi_{k,m}$	Set of subcarriers assigned to queue m of user k
ψ_k	Set of subcarriers assigned to user k
$N_{k,m}$	Number of subcarriers assigned to queue m of user k
λ	Lagrange multiplier associated with the power constraint
$\alpha_{k,m}$	Coding loss factor related to $\mathrm{BER}_{k,m}$ by adopting MQAM modulation
$\overline{LER_{k,m}}$	Average layer error rate for layer m of user k with equal power distribution across all subcarriers
$\beta_{k,m}$	Expected quality increment by increasing one unit of $LER_{k,m}$
$P_{k,max}$	Maximum allowable transmit power that can be allocated to a layer of user k

5.2 Overview of JPEG 2000

JPEG 2000 (for short, J2K) is a promising image coding standard developed by the ISO/IEC Joint Photographic Experts Group (JPEG) ("ISO/IEC 15444-1," 2000). Compared to its predecessor named JPEG, J2K uses subband-based discrete wavelet transform (DWT) rather than block-based discrete cosine transform (DCT). The reason behind this choice underlies the fact that DCT decomposes the source as a linear combination of disjoint blocks (e.g., 8x8-pixel blocks in JPEG), thus yielding a relatively low image data decorrelation (D. S. Taubman & Marcellin, 2002). Instead, DWT decomposes a whole image or large blocks into a number of subbands of horizontal, vertical and diagonal spatial frequencies. Combined with a more sophisticated entropy encoding mechanism, (Man et al., 2005) has verified that J2K consistently achieves 3 dB increase in PSNR versus JPEG under the same target bpp.

Recall that H.264 intra coding in previous chapter can be viewed as another alternative image coding method. The major difference between H.264 intra coding and J2K lies in the fact that H.264 intra coding relies on spatial prediction and DCT whereas J2K solely depends on DWT. Clearly, these two mechanisms are essentially different and such incompatibility is a testament to the robustness of the proposed BRA. Some features in the J2K especially those important to the aspects of this dissertation (from the communication perspective) are highlighted below. A detailed overview of the J2K codec is available in (Man et al., 2005; D. S. Taubman & Marcellin, 2002).

5.2.1 The J2K Syntax

An input image is separated into contiguous, non-overlapping tiles that are independently coded as depicted in figure 5.1. At the very beginning of the encoding process, a DC level shift is applied to each image component. For an image component

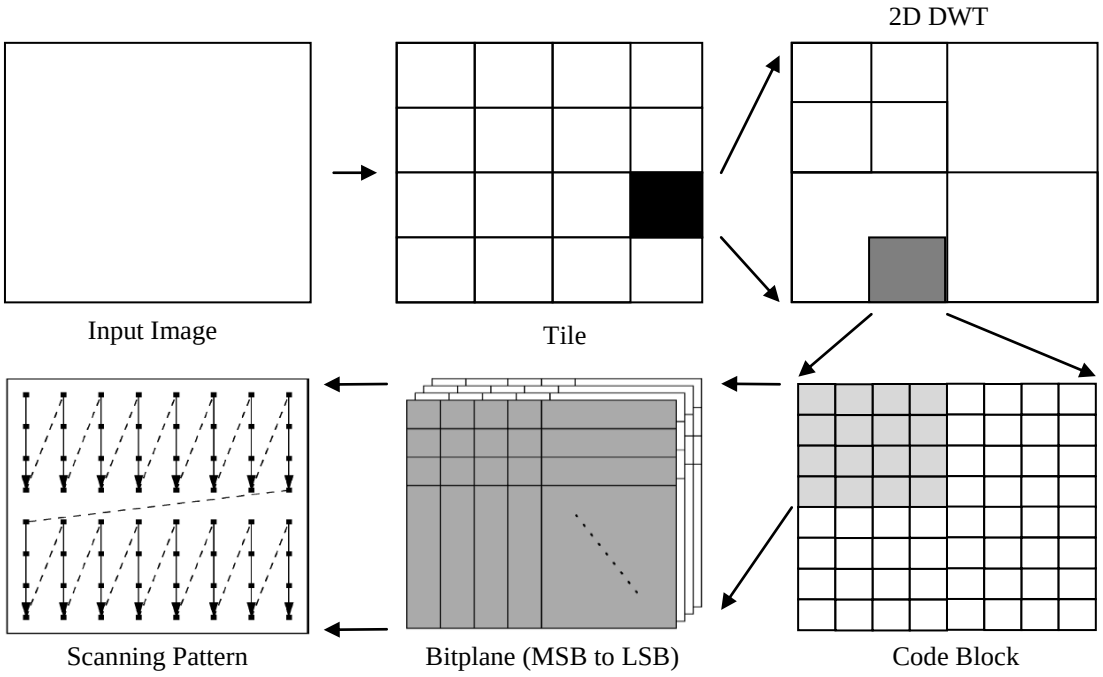


Figure 5.1 J2K syntax: low-level overview (Man, Docef, & Kossentini, 2005).

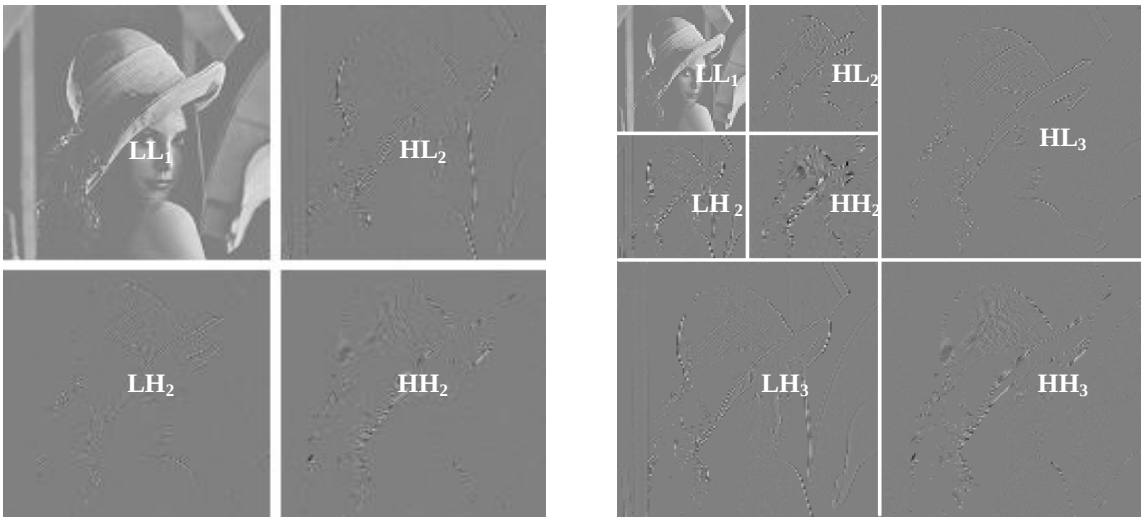
of bit depth B_d, the quantity $2B_d$-1 is subtracted from all the pixel values, in an attempt to retrieve a zero-mean distribution. This level shift will be offset at the decoder.

Each tile undergoes the same four main stages: 2D DWT, quantization, bitplane arithmetic coding and rate-distortion optimization. At the first stage, one-level 2D DWT decomposes the tile into four subbands: LL, LH, HL and HH. LL is the input tile, low-pass filtered in horizontal and vertical directions and downsampled by a factor of two; LH is low-pass filtered in horizontal direction and high-pass filtered in vertical direction and then downsampled by a factor of two, and so forth. This is achieved by applying a 1D wavelet transformation along the rows (or columns) of a 2D array and then the columns (or rows) of the resulting array.

J2K defines two sets of 1D wavelet transformation: 5/3 and 9/7. The former performs reversible integer-to-integer transformation via special rounding rules that can be used to achieve lossless coding. The latter undergoes real-to-real transformation that is irreversible due to finite precision floating-point architecture. Thus, it is used for lossy coding. Note that both wavelet transformations are implemented using a series of lifting

and scaling steps, which are not discussed in detail in this thesis. For an O-level 2D DWT, the same 2D DWT operation is iteratively applied to the $LL_{o=1}$ to create another four subbands for O times, producing $O + 1$ resolution levels as depicted in figure 5.2. Except that the lowest resolution level (O=1) has only LL_1, every resolution level consists of the remaining three subbands. Each subband which consists of wavelet coefficients is partitioned into rectangular blocks called code-blocks (CBs).

At the second stage, quantization of all wavelet coefficients for every CB is applied. J2K uses a uniform scalar quantizer with a dead zone around zero. Note that in lossless coding, those coefficients are not quantized, or they can be treated as quantized by a step size of one. At the third stage, the embedded block coding with optimized truncation (EBCOT) method proposed by (D. Taubman, 2000) is applied to each CB independently. Except that the most significant bitplane (MSB) has only one cleanup coding pass (CP), every bitplane possesses significance propagation (SP), magnitude refinement (MR) and cleanup coding passes. For each coding pass, the coefficients are scanned in groups of four rows whenever possible, starting at the top of the CB. Then,



(a) (b)

Figure 5.2 O-level 2D 5/3 DWT decomposition for a 256x256 *lena* grayscale image**.**
(a) $O = 1$. (b) $O = 2$.

each time a coefficient is to be encoded, a context is formed based on the currently coded values of neighboring coefficients and previous bitplane of the current coefficient. The context and the symbol (“0” or “1”) are fed to a binary adaptive arithmetic coder called the MQ coder in order to generate a bitstream. At the same time, the bit rate consumption and the reduction in distortion caused by this CP are computed and recorded for the purpose of bitstream truncation in the final stage.

At the final stage, the coding passes among all CBs are distributed across M_k layers using a post-compression rate-distortion (PCRD) optimization scheme subject to a target bit rate constraint. Each layer consists of a certain number of consecutive coding passes from each CB. Clearly, layer 1 is a base layer while layers $m > 1$ are enhancement layers. A concatenation of these layers outputs the J2K bitstream as shown in figure 5.3. The more consecutive coding passes of each CB is available at the decoder, the higher decoded image quality.

5.2.2 Error Resilience Features

J2K defines several error resilience features, particularly working at the entropy coding level, with the goal of detecting and isolating an error event. First, similar to the concept of slice in H.264, the decoding of each CB is carried out independently. This ensures

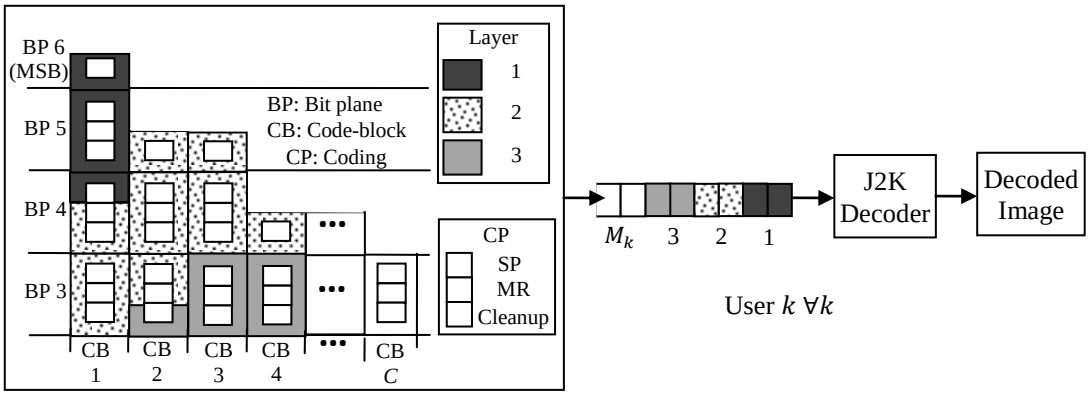


Figure 5.3 J2K scalable bitstreams.

that bit errors in one CB will not propagate to the rest. Second, a special arithmetic coded segmentation marker known as **SEGMARK** is inserted at the end of each bitplane. Upon detecting a corruption in the segmentation marker at the decoder side, an error is flagged in the current bitplane. Then, all the coding passes associated with the current and subsequent bitplanes are discarded. Third, a fine-granularity error detection method that operates at the CP level is to use **RESTART** in conjunction with **ERTERM** features. In **RESTART** mode, the MQ coder is restarted at the beginning of each CP. In this case, each CP has its own MQ codeword segment and there is no decoding synchronization issue among the CPs. In **ERTERM** mode**,** a predictable termination policy is employed for each MQ codeword segment at the encoder and thus it allows the decoders to detect errors by exploiting this property. With **RESTART-ERTERM**, only the affected CP and its following CPs are discarded. To conceal those discarded CPs, a simple method known as **ZERO FORCING** is to fill the missing bits of the wavelet coefficients with zero. More advanced concealment techniques can be found in (Zargari & Fatemi, 2005).

5.3 System Model

All J2K scalable bitstream are transmitted via an OFDMA system that as explained in section 3.2. Similar to the case in H.264, different layers from different users exhibits different contributions to the reconstructed quality $\Delta Q_{k,m}$ (defined later). Note that channel coding is not considered in this chapter as we have already examined its effect in chapter 4. Hence, it is worthwhile to investigate the potential gain of the BER-driven resource allocation under an uncoded OFDM system. Another reason for bypassing this option is that we intend to ensure fair comparisons as the benchmark schemes do not consider such settings. Unless specified otherwise, the assumptions and notations of the OFDMA system are the same as in the system considered in chapter 3.

During the decoding process, we assume the J2K decoder is able to detect and prevent errors from propagating through the entire bitstream (Moccagatta et al., 2000), with the help of error resilience features as discussed in section 5.2.2. That is, for each CB, the decoder can decode up to the coding passes in preceding layer if the current coding pass is corrupted. The missing coding passes are concealed with **ZERO FORCING**. We further assume that all the headers and markers used to prevent error propagation are transmitted error-free.

As the goal is to maximize the expected quality of all users $E[Q] = \sum_{k=1}^{K} E_k[Q]$ under the total power constraint P_{tot}, the optimization problem can be formulated as

$$\max_{\{\boldsymbol{\rho},\mathbf{s}\}} \sum_{k=1}^{K} E_k[Q]$$

s.t. (3.2), (3.6), (3.7),

$$\sum_{k=1}^{K} \sum_{m=1}^{M_k'} \sum_{n=1}^{N} \rho_{k,m,n} s_{k,m,n} \leq P_{tot}$$

$$\Gamma_{k,m,n} \in \{0, \dots, 2^b, \dots, 2^{b_{max}}\} \; \forall k, \forall m \text{ and } \forall n. \tag{5.1}$$

Note that M_k is the number of layers available at the BS while M_k' is the number of queues to be transmitted. Also, the last constraint represents the constraint of discrete modulation level as in chapter 4.

Now we will discuss the definitions of $R_{k,m}$, $\Delta Q_{k,m}$and $E_k[Q]$. Suppose there are a total of C CBs and CB c has $W_{k,c}^{max}$ coding passes for user k. In the PCRD optimization algorithm, the rate $\Delta r_{k,c,w}$ and the corresponding distortion reduction in terms of normalized MSE $\Delta d_{k,c,w}$ of the coding pass w in CB c are measured. Then, the received distortion $D_{k,m}$ for layers 1 to m and $R_{k,m}$ can be expressed as

$$D_{k,m} = D_{k,max} - \sum_{c=1}^{C} \sum_{w=1}^{W_{k,c}^{m}} \Delta d_{k,c,w} \tag{5.2}$$

$$R_{k,m} = \sum_{c=1}^{C} \sum_{w=W_{k,c}^{m-1}+1}^{W_{k,c}^{m}} \Delta r_{k,c,w}\ , \tag{5.3}$$

respectively. Here $D_{k,max}$ represents the distortion without decoding any coding pass for user k and $W_{k,c}^{m}$ is the total number of coding passes of CB c in the layers 1 to m,

$$0 \le W_{k,c}^{m-1} \le W_{k,c}^{m} \le W_{k,c}^{max}, \forall c \text{ and } 1 < m \le M_k. \tag{5.4}$$

which satisfies

Since $D_{k,max}$ is a constant, the term $\sum_{c=1}^{C} \sum_{w=1}^{W_{k,c}^{m}} \Delta d_{k,c,w}$ can be interpreted as accumulative layer quality contribution index for layer 1 to m

$$Q_{k,m} = \sum_{c=1}^{C} \sum_{w=1}^{W_{k,c}^{m}} \Delta d_{k,c,w} \text{ and } Q_{k,1} < Q_{k,\dots} < Q_{k,M_k}. \tag{5.5}$$

$Q_{k,m}$ can only be exploited if the previous $m-1$ layers have been correctly received owing to the progressive nature of the scalable bitstream. Let $LER_{k,m}$ be the layer error rate for layer m of user k. Then, $E_k[Q]$ can be written as

$$\begin{aligned} E_k[Q] = Q_{k,0} LER_{k,1} + \sum_{m=1}^{M_k-1} Q_{k,m} LER_{k,m+1} \prod_{f=1}^{m} \left(1 - LER_{k,f}\right) \\ + Q_{k,M_k} \prod_{f=1}^{M_k} \left(1 - LER_{k,f}\right) \end{aligned} \tag{5.6}$$

where $Q_{k,0} = 0$ as no information is available at the J2K decoder. Let $\Delta Q_{k,m}$=$Q_{k,m} - Q_{k,m-1}$ be the individual layer quality contribution index. Now, equation (5.6) can be further rewritten as given in (Cao, 2007)

$$E_k[Q] = \sum_{m=1}^{M_k} \prod_{f=1}^{m} (1 - LER_{k,f}) \Delta Q_{k,m} \tag{5.7}$$

Problem (5.1) is nontrivial as its objective function is a nonlinear function of $LER_{k,f}$. For this reason, we decompose this problem into three stages: dynamic target BER assignment, unequal subcarrier assignment, and bit and power allocation (algorithm 4.1).

5.4 Three-Level Framework for Scalable Bitstreams

At the first stage, we determine the number of transmit layers and the target BER of each selected layer simultaneously based on $\Delta Q_{k,m}$ and channel conditions. During poor channel conditions, the halt transmissions of less critical layers are necessary to avoid severe degradation of overall transmission reliability. At the second stage, we assign subcarriers to each layer based on the computed BERs from the first stage. Our strategy is to assign good subcarriers to important layers and *vice versa*. This make senses intuitively since less power will be consumed by the most important layer if it is transmitted by the best subcarrier as compared to that of an average subcarrier. Then the extra power budget benefitted from such idea can be allocated to the remaining layers to further improve the expected image quality. This conforms to the goal of problem (3.8), which maximizes power efficiency while transmitting certain number of bits. Hence, we adopt algorithm 3.1 to assign subcarriers to different layers. At the final stage, we employ frequency-selective power allocation rather than equal power allocation, following the conclusion from chapter 4.

5.4.1 Dynamic Target BER Assignment

At this stage, we assume average channel conditions without exploiting the frequency selectivity in the OFDMA system. More precisely, we consider individual subcarrier CNR $\gamma_{k,n}$ as average subcarrier CNR $E[ch_k]$,

$$\gamma_{k,n} = E[ch_k] \qquad \forall k \text{ and } \forall n. \tag{5.8}$$

Under this assumption, we aim to determine M_k' layers to be transmitted for user k and the target BER of each selected layer under the total power constraint P_{tot}. Both goals are achieved simultaneously by using the well-known *water-filling* theory. Three steps are required to accomplish the goal.

In the first step, we establish a relation $LER_{k,m} = f(s_{k,m,n})$ so that the objective function $E_k[Q]$ in equation (5.7) can be expressed in terms of transmit power. To obtain the relation, we assume the number of subcarriers assigned to layer m of user k $N_{k,m}$ is proportional to the size of layer $R_{k,m}$, which is formulated as

$$N_{k,m} = N\left(\frac{R_{k,m}}{\sum_{k=1}^{K}\sum_{m=1}^{M_k} R_{k,m}}\right) \quad \forall k \text{ and } \forall m. \tag{5.9}$$

Since we assume constant CNR across all subcarriers, the specific subcarrier assigned to each layer is not important at this point. Let $s_{k,m,n'}$ be the transmit power per subcarrier in $N_{k,m}$ subcarriers assigned to layer m of user k. Then, according to equation (3.3), we can write

$$R_{k,m} = N_{k,m}\log_2(1 + \alpha_{k,m}E[ch_k]s_{k,m,n'})(\tau/T_{ofdm}) \quad \forall k \text{ and } \forall m. \tag{5.10}$$

We now rewrite equation (5.10) as

$$BER_{k,m} = \frac{1}{5}\exp\left(\frac{-1.5E[ch_k]s_{k,m,n'}}{2^{(T_s\sum_{k=1}^{K}\sum_{m=1}^{M_k}R_{k,m}/\tau N)}-1}\right)\forall k \text{ and } \forall m. \quad (5.11)$$

We further assume all bits are independent and are equally protected in layer m in which $LER_{k,m}$ can be written as a function of $BER_{k,m}$

$$LER_{k,m} = 1-\left(1-BER_{k,m}\right)^{R_{k,m}} \forall k \text{ and } \forall m. \quad (5.12)$$

$$LER_k \approx R_{k,m}BER_{k,m} \quad \forall k \text{ and } \forall m. \quad (5.13)$$

By substituting equation (5.11) into (5.13), the relation $LER_{k,m} = f\left(s_{k,m,n}\right)$ is established. Although the objective function in problem (5.1) is now related to $s_{k,m,n}$, solving this problem turn out to be extremely difficult due to the nonlinearity nature of the function in equation (5.7). Therefore, we construct a linear $E_k[Q]$ function through a point-wise linear approximation (Atzori, 2003) in the next step to keep the optimizations computationally less intensive.

In the second step, we linearize equation (5.7) into

$$E_k[Q] \cong E_k[Q]\left(\overline{LER_{k,m}},\cdots,\overline{LER_{k,M}}\right)+\sum_{m=1}^{M_k}\frac{\partial E_k[Q]}{\partial LER_{k,m}}\left(LER_{k,m}-\overline{LER_{k,m}}\right)$$

$$\cong \eta_k-\sum_{m=1}^{M_k}\beta_{k,m}LER_{k,m} \quad \forall k, \quad (5.14)$$

where $\overline{LER_{k,m}}$ is average layer error rate for layer m of user k with equal power distribution across all subcarriers. In particular, it is obtained by substituting $s_{k,m,n'} = P_{tot}/N$ into equations (5.11) and (5.12). η_k is a constant while $\frac{\partial E_k[Q]}{\partial LER_{k,m}}$ and positive constant $\beta_{k,m}$ are given by

$$\frac{\partial E_k[Q]}{\partial LER_{k,m}} = -\beta_{k,m} = -\sum_{g=m}^{M_k} \Delta Q_{k,g} . \prod_{\substack{f=1 \\ f\neq m}}^{g} \left(1 - \overline{LER_{k,f}}\right) \ \forall k \text{ and } \forall m \,, \tag{5.15}$$

respectively. $\frac{\partial E_k[Q]}{\partial LER_{k,m}}$ represents the expected quality increment by decreasing (negative sign) one unit of $LER_{k,m}$. Obviously, the hierarchical structure of the scalable J2K bitstream implies $\beta_{k,1} > \beta_{k,2} > \cdots > \beta_{k,M_k} > 0$.

In the third step, we maximizes $E[Q]$ subject to the power constraint which is formulated as

$$\max_{s_{k,m,n'}} \sum_{k=1}^{K} E_k[Q] \text{ in (5.14)}$$

s.t.
$$\sum_{k=1}^{K}\sum_{m=1}^{M_k} s_{k,m,n'} = \frac{P_{tot}\sum_{k=1}^{K} M_k}{N} \ . \tag{5.16}$$

Note that the reason why P_{tot} is multiplied by $(\sum_{k=1}^{K} M_k \,/N)$ is we assign a group of subcarriers to each layer in equation (5.9). The problem (5.16) is equivalent to finding the maximum of the following Lagrange function

$$\begin{aligned}\mathcal{L}\left(\{s_{k,m,n'}\},\lambda\right) = \sum_{k=1}^{K}\eta_k - \sum_{k=1}^{K}\sum_{m=1}^{M_k}\beta_{k,m}\frac{R_{k,m}}{5}\exp\left(\frac{-1.5E[ch_k]s_{k,m,n'}}{2^{(T_s\sum_{k=1}^{K}\sum_{m=1}^{M_k}R_{k,m}/\tau N)}-1}\right) \\ -\lambda\left\{\sum_{k=1}^{K}\sum_{m=1}^{M_k} s_{k,m,n'} - \frac{P_{tot}\sum_{k=1}^{K}M_k}{N}\right\}\end{aligned} \tag{5.17}$$

By differentiating equation (5.17) with respect to $s_{k,m,n'}$ and setting each derivative to zero, the optimal $s^*_{k,m,n'}$ can be obtained as

$$s^*_{k,m,n'} = \left[ln(A_{k,m}/\lambda).\theta_k\right]^{+} \forall k \text{ and } \forall m, \tag{5.18}$$

$$A_{k,m} = \left(\frac{\beta_{k,m} R_{k,m}}{5}\right)\left(\frac{1.5E[ch_k]}{2^{(T_s \sum_{k=1}^{K} \sum_{m=1}^{M_k} R_{k,m}/\tau N)} - 1}\right) \quad \forall k \text{ and } \forall m, \tag{5.19}$$

where

$$\theta_k = \frac{(2^{(T_s \sum_{k=1}^{K} \sum_{m=1}^{M_k} R_{k,m}/\tau N)} - 1)}{1.5E[ch_k]} \quad \forall k \tag{5.20}$$

and $[x]^+ \triangleq \max\{x, 0\}$.To find the optimal λ_0 that maximizes$\sum_{k=1}^{K} E_k[Q]$, the update of λ can be done using the bisection method until the sum power converges to $(P_{tot} \sum_{k=1}^{K} M_k)/N$.

Solution (5.18) has same flavor as *water-filling* theorem (Cover & Thomas, 1991). However, distinguishing from the work that *water-fills* over frequency (Jang & Lee, 2003), solution (5.18) pours transmit power over all layers depending on $A_{k,m}$: importance of layer indicated by $\beta_{k,m}$, size of each layer $R_{k,m}$, average channel conditions $E[ch_k]$ and allocation period τ. The larger the $A_{k,m}$, the more transmit power is allocated to layer m. If $A_{k,m}$ drops below λ_0 , solution (5.18) suggests that the layer is not transmitted since $s^*_{k,m,n'} = 0$. This indicates that LI layers with smaller $A_{k,m}$ (due to smaller $\beta_{k,m}$) tend to be halted transmitting as compared to MI layers. This makes sense intuitively because when current average channel condition of user k is poor ($E[ch_k]$ small), a better option is not to transmit LI layers so that large amounts of the transmit power are used to maintain the overall transmit reliability of MI layers.

Besides that, the larger the $R_{k,m}$, the more transmit power is allocated layer m. This is intuitive since larger amount of data bits requires more resources. Also, LI layers will not be transmitted when allocation period τ is short. τ can be interpreted as frame delivery deadline in the context of video transmission. If the frame delivery deadline is tight, any video packet that cannot meet their deadlines are dropped before the transmission.

The solution to problem (5.16) is guaranteed to converge to the global maximum that satisfies the power constraint because the *Hessian* matrix of $\mathcal{L}(\{s_{k,m,n'}\},\lambda)$ is negative definite for all possible **s.** Once $s^*_{k,m,n'}$ satisfying the power constraint in problem (5.16) is found, $BER_{k,m}$ is computed using equation (5.11) and will be utilized in the next stage. Recall that the layers with zero $s^*_{k,m,n'}$ are not transmitted. Hence, only M'_k layers of user k will be transmitted.

Figure 5.4 elaborates more clearly the *water-filling* concept of solution (5.18). In this figure, there are $K = 1$, $M_{k=1} = 5$ layers available at the BS. Based on λ_0, only $M'_{k=1} = 3$ layers are transmitted since their values of $(\ln A_{k,m} - \ln \lambda_0)$ are positive, which satisfy constraint (3.2).

Since $A_{k,1} \gg A_{k,m} \gg A_{k,M_k}$ (based on test images in our experiment), directly applying equation (5.18) will cause overprotection of MI layers especially when channel conditions are good (Hamzaoui et al., 2005). In other words, MI layers are always water-filled with excessive amounts of transmit power. To address this issue, we first set a minimum target BER BER_{min} to determine the maximum allowable transmit power that can be allocated to a layer of user k $P_{k,max}$which is obtained from equation (5.11) with $BER_{k,m} = BER_{min}$,

$$P_{k,max} = \frac{-\ln(5BER_{min}) \cdot \left(2^{(T_s \sum_{k=1}^{K} \sum_{m=1}^{M} R_{k,m}/\tau N)} - 1\right)}{1.5E[ch_k]} \quad \forall k \tag{5.21}$$

If the amount of transmit power allocated to a layer of user k exceeds $P_{k,max}$, only $P_{k,max}$ is allocated to that layer and the remaining power will be allocated to other layers. In this way, less important layers obtain more transmit power while ensuring more important layers are well-protected against channel errors. The complete steps are given in algorithm 5.1.

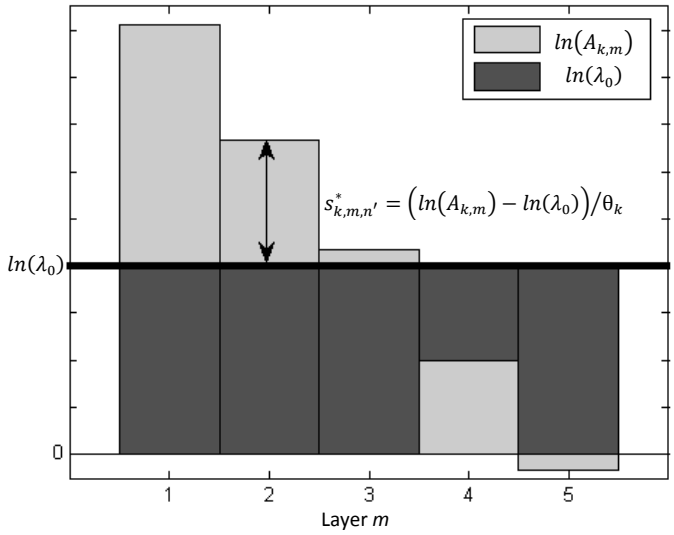


Figure 5.4 *Water-filling* concept from solution (5.18)

Table 5.2 Algorithm 5.1 Dynamic target BER assignment.

1: Initialization: Choose two points a_1 and a_2 such that the optimal point λ_0 lies in the interval (a_1, a_2). Also choose $0 < \epsilon \ll 1$ as the stopping criterion. Set iteration index $i = 1$.

2: Set BER_{min}, compute $P_{k,max}$ using equation (5.21) $\forall k$ and set $P_{constraint} = \frac{P_{tot}\sum_{k=1}^{K} M_k}{N}$. Let $N_{k,max}$ denote the number of layers that has been assigned with $P_{k,max}$ and set $N_{k,max} = 0$. Set $s^*_{k,m,n'} = P_{k,max}$ and $M'_k = 0$.

3: **while** $s^*_{k,1+N_{k,max},n'} \geq P_{k,max}$ **AND** $N_{k,max} \neq M_k - 1$

4: **for** k= 1 to K

5: **If** $i \neq 1$ **AND** $s^*_{k,1+N_{max},n'} = P_{k,max}$, set $N_{k,max} = N_{k,max} + 1$ and $P_{constraint}$= $P_{constraint} - P_{k,max}$.

6**:** **end for**.

7: Set $\lambda_{low} = a_1, \lambda_{high} = a_2$.

8: **repeat**

9: set $\lambda = (\lambda_{low} + \lambda_{high})/2$.

10: **for** k= 1 to K; **for** m= $(N_{k,max} + 1)$ to M_k

11: Compute $s^*_{k,m,n'}$ using equation (5.18).

12: **end for; end for.**

13: **If** $\sum_{k=1}^{K} \sum_{m=1+N_{k,max}}^{M_k} s^*_{k,m,n'} > P_{constraint}$, set $\lambda_{low} = \lambda$; E**lse**, set $\lambda_{high} = \lambda$.

14: **until** $\left| \sum_{k=1}^{K} \sum_{m=1+N_{k,max}}^{M_k} s^*_{k,m,n'} - P_{constraint} \right| < \epsilon$

15: **If** $s^*_{k,1+N_{k,max},n'} > P_{k,max}$, set $s^*_{k,1+N_{max},n'} = P_{max}$. $\forall k$

16: $i = i + 1$.

17: **end while**

18: **for** k= 1 to K; **for** m= 1 to M_k

19: **If** $s^*_{k,m,n'} > 0$, calculate $BER_{k,m}$ using equation (5.11). $M'_k = M'_k + 1$.

20: **end for; end for.**

No computation is needed in obtaining the content information since the J2K encoding algorithm already supplies this information (Banister, Belzer, & Fischer, 2002). Line 3 requires at most a complexity of $\mathcal{O}(\sum_{k=1}^{K} M_k)$. Let I_2 be the number of iterations needed using bisection search to reach the conditions in line 14. In each iteration, computing $s^*_{k,m,n'}$ in line 11 needs at most a complexity of $\mathcal{O}(\sum_{k=1}^{K} M_k)$. Then, computing $BER_{k,m}$ in line 19 requires a complexity of $\mathcal{O}(\sum_{k=1}^{K} M'_k)$. Therefore, the worst-case

computational complexity of algorithm 5.1 is $\mathcal{O}(I_2(\sum_{k=1}^{K} M_k)^2 + \sum_{k=1}^{K} M_k')$. In our experiment, I_2 is around 125.

5.4.2 Unequal Subcarrier Assignment

At this stage, we assign subcarriers to M_k' layers of user k based on the computed $BER_{k,m}$ from previous stage. Such strategy can be realized with algorithm 3.1. However, this algorithm is still fairly complex to implement due to the iterative computation. In order to reduce the computational complexity, we develop a simple suboptimal unequal subcarrier assignment that consists of two steps. The first step tends to assign a subcarrier to the user with the best channel gain on that subcarrier while maintaining the fairness across different users. The second step utilizes *proposition 3.1*, which assigns good subcarriers to important layers, and vice versa. Note that the second step is performed in isolation for each user. The algorithm can be described as

Table 5.3 Algorithm 5.2 suboptimal unequal subcarrier assignment.

1: Set $\psi_k' = \emptyset$ and calculate the average number of subcarriers assigned to user k l_k with $l_k = N\left(\sum_{m=1}^{M_k'} R_{k,m} / \sum_{k=1}^{K}\sum_{m=1}^{M_k'} R_{k,m}\right) \forall k$.

2:**for** $n = 1$ to N

3: $k' \leftarrow \arg\max_k \gamma_{k,n}$,

4: **If** $l_k < |\psi_k'|$, $\psi_{k'}' = \psi_{k'}' \cup \{n\}$; **Else**, exclude k' and go to line 3.

5: **end for.**

6: **for** $k = 1$ to K

7: Rearrange the subcarrier indexes in ψ_k' in descending order in terms of subcarrier

CNR $\gamma_{k,n}$ (e.g., $\gamma_{k,1} \geq \cdots \geq \gamma_{k,N}$, set $\psi_{k,m} = \emptyset$ for $m = 1,2 \ldots, M_k'$ and $|\psi_{k,0}| = 0$.

8: **for** $m = 1,2 \ldots, M_k'$

9: Calculate the number of subcarriers assigned to layer m of user k $|\psi_{k,m}|$ with $|\psi_{k,m}| = |\psi_k'| \left(R_{k,m} / \sum_{m=1}^{M_k'} R_{k,m} \right) \forall m$

10: set $\psi_{k,m} = \psi_{k,m} \cup \{ \sum_{t=0}^{m-1} |\psi_{k,t}| + 1, \ldots, \sum_{t=0}^{m} |\psi_{k,t}| \}$

11: **end for; end for.**

In line 1, it needs $\mathcal{O}(K)$ operations. In line 3, it requires a complexity of $\mathcal{O}(KN)$. If without line 4, a user may not be allocated any subcarrier if the user has no best subcarrier. In line 7, we need to sort the subcarrier CNR for each user, leading to a complexity of $\mathcal{O}(\log_2|\psi_k'|)$. So the total computational complexity of algorithm 5.2 is $\mathcal{O}\big(K + KN + K(\log_2|\psi_k'| + M_k')\big)$, which is much simpler as compared to the that of algorithm 3.1 because no iteration is needed.

5.5 Simulation Results

In the simulations, several standard images (256 × 256 gray-scale, B_d=8 bits/pixels) were used. The images were encoded using JasPer 1.900.1 ("JasPer software reference manual," 2002) and processed with 5/3 reversible 2D DWT with one-level ($O = 1$) decomposition before dividing the images into code-blocks of 16 × 16. Then, the images were compressed with a target bit rate of 2.886 bits/pixel and formed into five layers ($L = 5$). The size of each layer was set to be the same, thereby contributing to a rate of 0.577 bpp.

For the OFDMA system, the total bandwidth $B_T = 1$ MHz , the total transmit power $P_{tot} = 1$ W, and the number of subcarriers $N = 128$. The allocation period τ was set in such a way that the total transmit rate per OFDM symbol equals to 512 bits/symbol (or equivalently, an average of four bits/subcarrier). For example, the five layers of *barbara* consists of (2.886*256*256) bits need $\tau = 367 * T_{ofdm}$. Recall that we assume $\gamma_{k,n}$ remains constant within each allocation period and $367 * T_{ofdm}$ might not be short enough so that the channel gain is stable within a transmit time interval. Nevertheless, if the allocation period τ is set shorter than predetermined one, algorithm 5.1 takes into account the impact of shorten τ value and might transmit fewer layers. Note that we set $BER_{min} = 10^{-6}$ in algorithm 5.1. For the discrete modulation, we assume five different MQAM constellations with gray bit mapping are available, that is $\Gamma \in \{0, 4, 16, 64, 256, 1024\}$ corresponding to 0, 2, 4, 6, 8 and 10 bits per symbol. For each subcarrier n, individual subcarrier $SNR_{k,n} = \gamma_{k,n} s_{k,m,n}$ is computed and channel errors are injected to the modulated bitstream according to $SNR_{k,n}$. Then, simulated BER of each layer is computed. The rest of the wireless system parameters are the same as in section 3.4.2 except that channel coding part is ruled out. We denote average SNR (ASNR) of user k by $(P_{tot} E[ch_k]/N)$.

Three images *barbara*, *boat* and *lena* were selected as the test images. We compare five algorithms as follows.

1. Optimal (opt) BRA-SB: It is the algorithm that executes algorithms 5.1, 3.1 and 4.1.

2. Suboptimal (sub) BRA-SB: It is the algorithm that executes algorithms 5.1, 5.2 and 4.1.

3. Static BRA-SB: It is the algorithm that executes algorithms 3.1 and 4.1 with all five layers transmitting. The target BER of each layer is static regardless of channel conditions and layer importance, which is similar to (Su et al., 2006; Yen et al.,

2010; Zhu, 2012). We arbitrarily set $\mathrm{BER}_{\mathrm{k},1} = 10^{-6}, \mathrm{BER}_{\mathrm{k},2} = 10^{-5}, \mathrm{BER}_{\mathrm{k},3} = 10^{-4}$, and $\mathrm{BER}_{\mathrm{k},4} = \mathrm{BER}_{\mathrm{k},5} = 10^{-3}$.

4. Without-BRA-SB-all (WBRA-SB-all): It is the algorithm that transmits all five layers as a single bitstream without considering the BER diversity. Only algorithm 4.1 with $\mathrm{BER}_{\mathrm{k,m}} = 10^{-6}$ $\forall$m is executed. Therefore, the goal is to maximize the total transmit rate, which is similar to (S. S. Tan et al., 2008) during the power allocation.

5. Without-BRA-SB (WBRA-SB): It is identical to WBRA-SB-all with the exception that the number of transmit layers is the same as that of opt BRA-SB and sub BRA-SB.

Figures 5.5 to 5.7 show the results in the one-user case, while figure 5.8 depicts the results in the two-user case. For the two-user case, the experiments were repeated in two scenarios in order to verify the robustness of BRA-SB algorithm. The former assume that there is no path loss difference between the two users while the latter corresponds to the existence of large-scale fading. Both scenario results are plotted in figure 5.8 (a) and (b), respectively.

Figure 5.5 highlights the PSNR versus ASNR curves for *barbara* by using different algorithms in the one-user case. Note that the term "(1-8)=1" indicates the number of transmit layers (1) under the ASNR range (1-8), and so forth. It can be observed that the number of transmit layers increases with ASNR for opt BRA-SB, sub BRA-SB and WBRA-SB. The performance of sub BRA-SB is very close to that of opt BRA-SB. At all points, the suboptimal algorithm achieves at least 90% of the optimal performance while offering a significant computational advantage. Additionally, opt BRA-SB and sub BRA-SB significantly outperform static BRA-SB and WBRA-SB-all in terms of PSNR. At low ASNR range spanning from 1 to 15 dB, the PSNR gains for Static BRA-

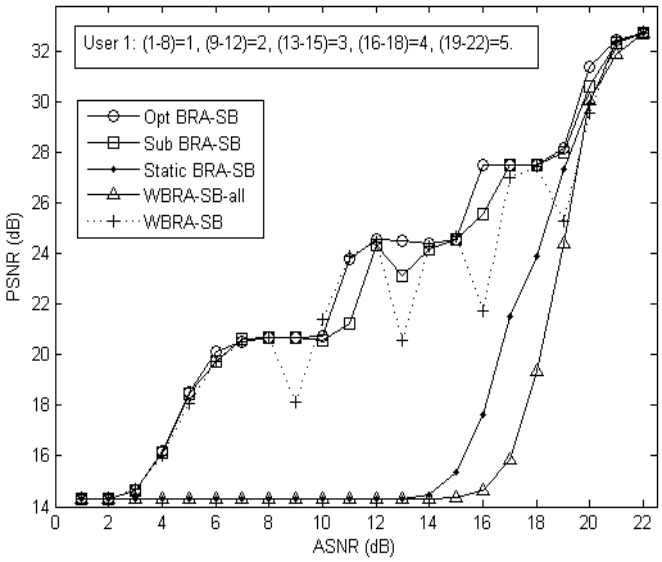


Figure 5.5 PSNR versus ASNR for *barbara* in one-user case.

SB and WBRA-SB-all are almost negligible (i.e., less than 1 dB). Meanwhile, at high ASNR range spanning from 17 to 20 dB, opt BRA-SB is 1.4-11.6 dB better than WBRA-SB-all. The advantage is contributed by our transmission strategy that prioritizes the protection of more important layers. On the contrary, WBRA-SB-all does not distinguish the importance of each layer and renders more important layers exposing to much higher BER than those in BRA-SB. At high ASNR region (over 20 dB), the superiorities of BRA-SB-like schemes weaken as good channel conditions allow whole bitstream to be transmitted securely.

To obtain a better insight for the performance of the five schemes above, figure 5.6 depicts the simulated BERs of opt BRA-SB, static BRA-SB, WBRA-SB-all and WBRA-SB. For opt BRA-SB and static BRA-SB, the simulated BERs of individual layers are plotted. The only exception is for Static BRA-SB, the average BER ABER) of layers 4 and 5 is shown since both layers share almost same results. Likewise, only ABER of all transmit layers are shown for WBRA-SB-like schemes. Note that the results of sub BRA-SB are not plotted in this figure for ease of illustration since its performance exhibits similar trend as opt BRA-SB with relatively higher BERs of individual layers. It can be seen that the BER of opt BRA-SB-L1 is much lower than

that of the rest layers, static BRA-SB and WBRA-SB-like schemes since BRA-SB gives the highest protection to the most important layer.

Besides that, although the BER of opt BRA-SB-L2 is much higher than the ABER of WBRA-SB at 9 dB ASNR, opt BRA-SB is 2.5 dB better than WBRA-SB in terms of PSNR. This implies that the image quality is very much determined by the location of bit errors instead of the amount of errors. There are four peaks for WBRA-SB at ASNR=9, 13, 16 and 19 dB respectively which explains the fluctuating PSNR curve for WBRA-SB in figure 5.5. These peaks are caused by the sudden increase in total number of transmit layers which drastically reduces the overall transmit reliability. Therefore, there is a high possibility of MI layers getting corrupted by channel errors. For static BRA-SB, the gaps in simulated BER between the five bitstreams are due to the predetermined target BERs.

Figure 5.7 visualizes the quality gain of BRA-SB-like schemes compared to the alternatives for *barbara* at ASNR =16 dB. It can be noticed that despite static BRA-SB

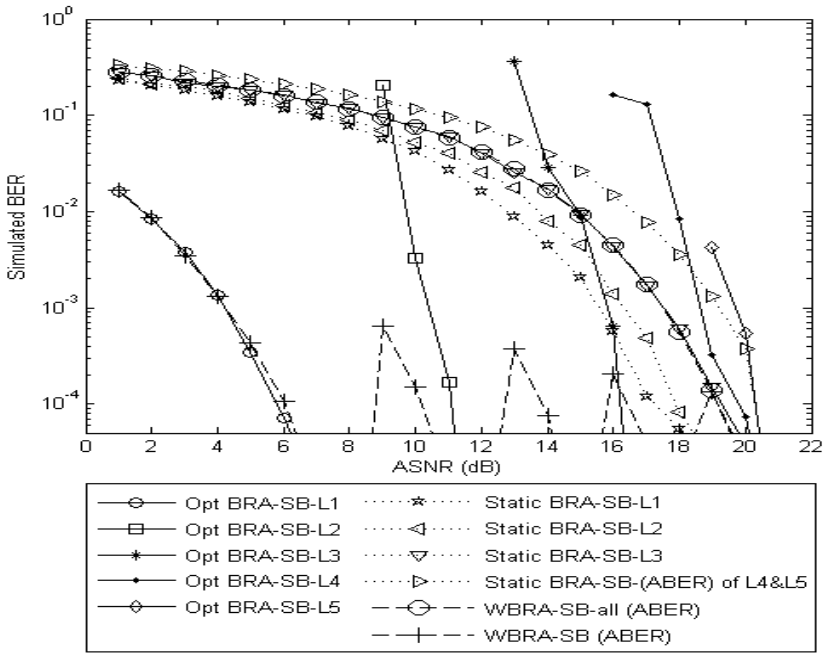


Figure 5.6 Simulated BER versus ASNR for *barbara* in one-user case.

and WBRA-SB-all transmit same amount of image data, the image for WBRA-SB-all is completely incomprehensible as compared to Static BRA-SB since the base layer for static BRA-SB is corrupted by channel noise.

Figure 5.8 (a) depicts the average PSNR across two users (*boat* and *lena*) versus the ASNR of both users. The overall performance of two-user case is quite similar to that of the one-user case, where opt BRA-SB outperforms the alternatives for most of the ASNR range. Again, the PSNR gap between opt BRA-SB and sub BRA-SB is very small. At ASNR=10 and 14 dB, the PSNR for WBRA-SB is slightly higher than that for opt BRA-SB due to the overprotection of more important layers. This renders the remaining layers ill-protected. Nevertheless, WBRA-SB achieves inferior overall results compared with BRA-SB-like methods due to its fluctuations in terms of PSNR over the ASNR range. The PSNR gain for opt BRA-SB over WBRA-SB-all is as high as 13.7 dB, which is larger than that in the one-user case owing to the multiuser diversity gain.

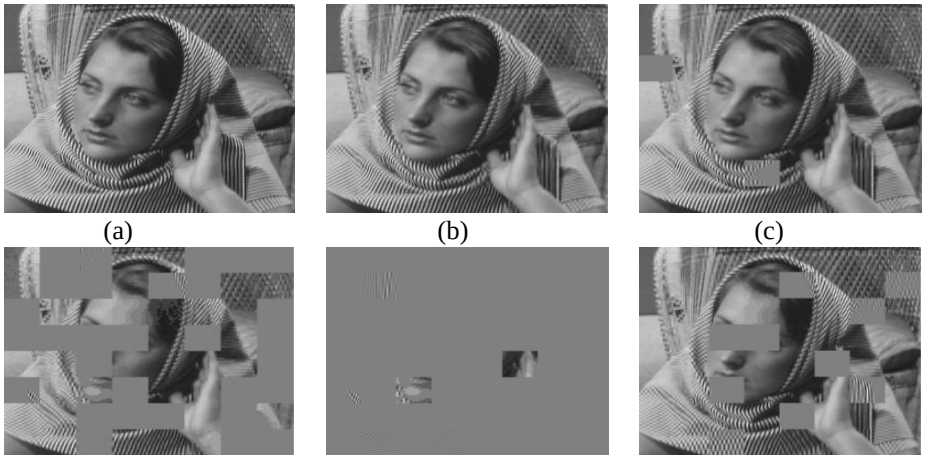


Figure 5.7. *Barbara* results for different algorithms at 16 dB ASNR.
(a) Original uncompressed image. (b) Opt BRA-SB. (c) Sub BRA-SB.
(d) Static BRA-SB. (e) WBRA-SB-all. (f) WBRA-SB.

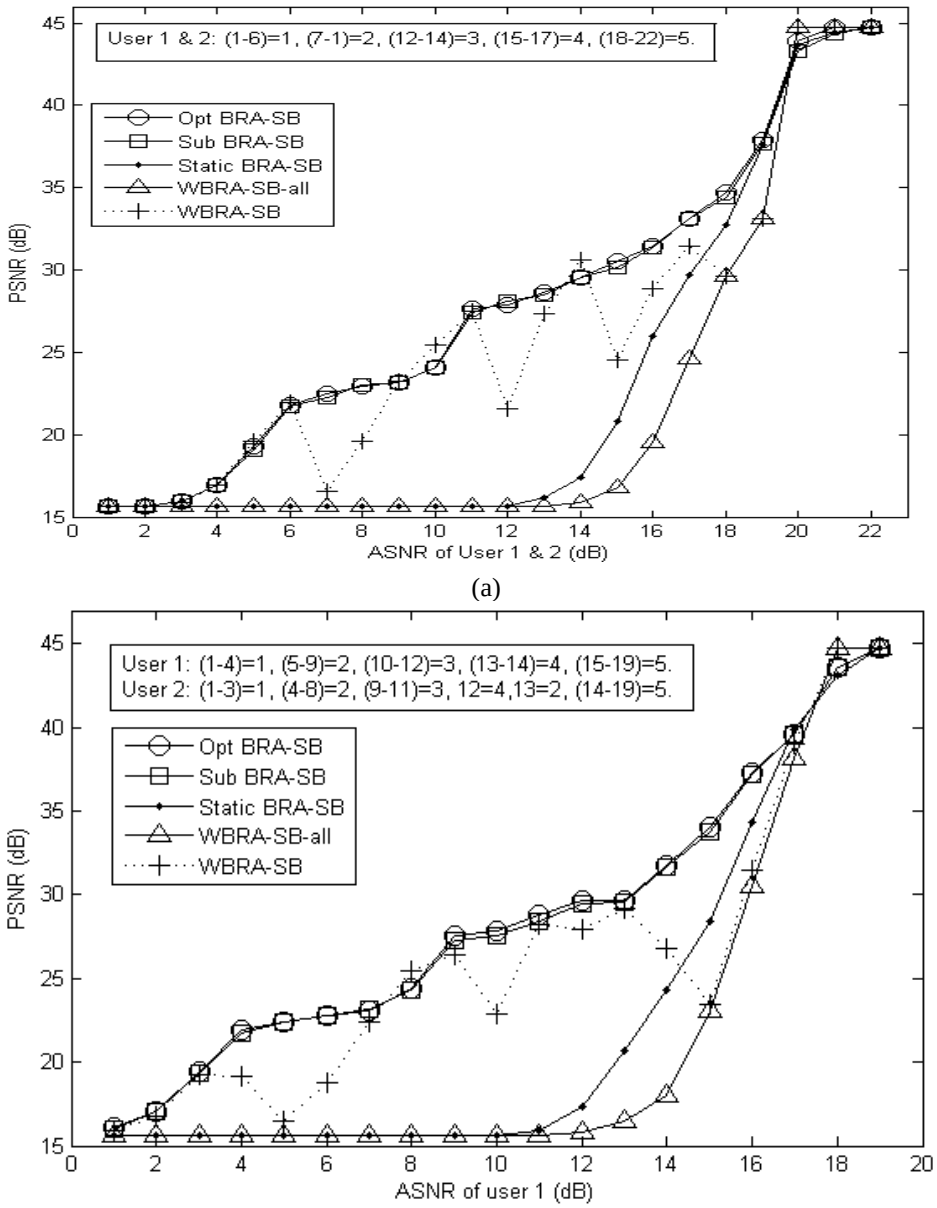


Figure 5.8 Average PSNR across two users versus ASNR of user 1.
(a) Equal average channel gain: $\boldsymbol{E[ch_1] = E[ch_2]}$. (b) User 2 has better average channel condition: $\boldsymbol{E[ch_2]/E[ch_1] = 5}$ dB. user 1: *boat* and user 2: *lena.*

Figure 5.8(b) shows the average PSNR across two users versus the ASNR of user 1, where the average channel CNR of user 2 $E[ch_2]$ is 5 dB higher than $E[ch_1]$. It can be seen that the number of transmit layers no longer necessarily increases with ASNR. At ASNR=13 dB, the number of transmit layers for user 2 reduces to two from four while the number of transmit layers for user 1 increases from three to four. The reason is that transmitting four layers of *boat* incurs much better image quality than that for *lena* (i.e., *boat* and *lena* : 36.2425 and 35.6892 dB in our experiment). Since user 2 with better

channel condition consumes less transmit power to transmit certain data, algorithm 5.1 executes to transmit fewer layers of *lena* so that more transmit power can be allocated to maximize the expected quality of user 1.

5.6 Summary

In this chapter, we have proposed a BER-driven resource allocation scheme for use in a scalable J2K framework. A *water-filling* algorithm has been developed to optimize the number of layers to be transmitted and to compute the target BER of the selected layer simultaneously, depending on the current channel conditions and the importance of each layer. Bit-level simulation results show that the proposed scheme accrues significant PSNR quality improvement over BER-unaware algorithms (i.e., WBRA-SB-all and WBRA-SB). The highest gains are 11.6 dB and 13.7 dB for the one-user and two-user cases, respectively. It is also demonstrated that that a suboptimal unequal subcarrier assignment, which reduces the complexity from $\mathcal{O}(I_1 \sum_{k=1}^{K} M_k' N)$ to $\mathcal{O}\big(K + KN + K(\log_2|\psi_k'| + M_k')\big)$, can still provide excellent results.

CHAPTER 6

Conclusions and Future Works

6.1 Conclusions

This thesis is dedicated to resource-allocation techniques for transmission of pre-encoded multimedia bitstreams over downlink OFDMA systems. The first chapter discusses the motivations behind the use of cross-layer resource allocation for multimedia communication and the challenges they confront, that have paved the way for the proposed BRA methods. The second chapter provides some background on the entire wireless multimedia communication system that leads to a realistic simulation setup. Subsequent chapters demonstrate that the utilization of multiple target BERs indeed opens the possibility of increased system performance, for three types of multimedia bitstream: generic, nonscalable and scalable.

For the generic case, we have shown that three major classes of resource allocation scheme with the consideration of multiple target BERs yield remarkable improvements over their counterparts in terms of spectral efficiency and power efficiency. The former is meaningful as the improved throughput can simultaneously support a larger number of data-intensive multimedia services. Equally important, the latter can turn into reduced energy consumption which is the core and active subject of green communications. Particular attention should be devoted to the sum-power minimization problem where we have theoretically proven that important multimedia bitstreams should be assigned subcarriers with good channel conditions in order to have better protection levels. In contrast, conventional works adopt this common assumption without any mathematical reason.

For both nonscalable and scalable bitstreams, we have demonstrated that the theoretical performance gains obtained in the generic case can be successfully translated into

improved realistic multimedia qualities. This is contributed to the key principle that optimizes the target BER of each multimedia packet as a function of the resource constraints, the channel conditions and the importance level of packet. The realistic performances take into account the real image and video codecs as well as practical modulation and coding schemes. With regard to the nonscalable case, simple and complex source-level error resilience methods have been considered. For the same transmitted packets, we have shown that BRA for nonscalable bitstreams significantly outperforms the state-of-the-art method that treats all packets with the same target BER, by up to 7.9 dB PSNR. On the other hand, the values of BRA for scalable bitstreams have been displayed as compared to the conventional BER-unaware methods. The highest gains are 11.6 dB and 13.7 dB PSNR for the one-user and two-user cases, respectively.

The relatively higher gain found in the scalable multimedia case is attributed to the fact that the decoding dependency of scalable bitstreams is much stronger than that of nonscalable one. In that case, the UEP offered by the BRA scheme could be more effective. This also contributes to the reason why we have adopted the expected quality instead of deterministic quality as the objective function for the scalable bitstream case. The mathematical derivations associated with the objective function together with the resource constraints, have culminated in a new *water-filling* concept. This technique pours transmission power over all packets as a function of the importance of packet, the length of packet, and the channel conditions. Furthermore, we have investigated the suitability of equal power allocation which is the common assumption in the OFDM literature. From the multimedia performance perspective, we have recommended the use of frequency-selective power allocation as equal power allocation incurs a substantial quality loss of 4.9 dB PSNR.

Since multimedia applications are delay-sensitive, lightweight algorithms that ease the computational burdens at the base station are preferred for implementations. In this regard, we have suggested two low-complexity suboptimal methods: one for estimation of video packet-level importance in chapter 4 and another one for unequal subcarrier assignment in chapter 5. We have revealed that both methods still hold the promise of enhanced multimedia quality.

6.2 Future Works

A number of future directions naturally grow out of the work developed in this thesis. Specifically, we are interested in the following studies.

The more ambitious future data traffic requirements are stimulating more innovative technology solutions. One promising candidate is multiuser multiple-input-multiple-output (MU-MIMO) systems, where a base station is communicating simultaneously to multiple users while both parties are equipped with multiple antennas. In fact, this technology is one of the key features in the Long Term Evolution-Advanced (LTE-A) (Lim, Yoo, Clerckx, Lee, & Shim, 2013), performance of which is still an open and active issue. As in single-input-single-output (SISO) systems (investigated in this work), resource allocation continues to play a critical role that boosts the performance of MU-MIMO systems (Capozzi, Piro, Grieco, Boggia, & Camarda, 2013). Thus, the proposed BRA schemes can be very beneficial in the MIMO framework, where the problem of resource allocation in a multiuser MIMO-OFDM system can be similarly formulated but more challenging due to more optimization variables associated with multiple antennas. Overall, it is interesting to see how much performance gains (especially the received multimedia quality) can be extracted in such framework.

Another direction that is worth looking at is related to practicability issue. This instantly links to the computational complexity issue. Although two simple techniques have been

proposed in this work, further investigation should be conducted to achieve the best tradeoff between performance and complexity. This effort should be greatly appreciated, concerning that the controlling resources in future wireless communication systems grow larger such as MIMO. Besides that, this thesis assumes that the CSI is perfect at the transmitter side. However, in real scenarios, the accuracy of channel information is strongly dependent on channel estimation error, feedback delay and the transmission reliability of the information (Z. Chang, Ristaniemi, & Niu, 2014). Therefore, the impact of imperfect CSI on the resource allocation schemes should be examined. While extensive similar works have been carried out in the past, most of them do not consider the impact of partial CSI from the multimedia performance perspective.

Another key area of research is uplink multimedia transmission, which is crucial for live coverage of sport events and camera surveillance systems. Similar to downlink, the objective of resource allocation for multiuser uplink transmission is to optimize the multimedia performance except for the three additional concerns. First, mobile stations instead of the base station act as transmitters this time. Therefore, the transmission power of each individual video user must be bounded in order to minimize inter-cell interference. Second, most uplink multimedia applications will need real-time multimedia encoding at the client side. In this case, joint source-channel coding technique can be a promising solution. Third, the multimedia content information cannot be immediately accessed at the base station as the encoded multimedia packets are now located at their corresponding mobile clients. Thus, effective content-specific information gathering methods should be developed so that the optimization of multimedia transmission continues to run in a centralized fashion. In this regard, the devised BRA methods can offer a basis for overcoming the challenges related to the allocation of resources among users such that high quality of multimedia services are provisioned across multiple flows. This will ultimately lead to content-adaptive (not

only content-aware) BER-driven resource-allocation strategy for multimedia transmission, which is left for future study of uplink multimedia transmission.

APPENDIX A

Based on the definition $G_{k,m,n}(s_{k,m,n}) \triangleq (s_{k,m,n} - \mu_{k,m} r_{k,m,n})$, $G_{1,m,n}(\bar{s}_{1,m,n}) \triangleq \bar{s}_{1,m,n} - \mu_{1,m}\{\log_2(1 + \alpha_{1,m}\gamma_{1,n}\bar{s}_{1,m,n})(\tau/T_{ofdm})\}$. As we intend to establish the relationship between $G_{1,m,n}(\bar{s}_{1,m,n})$ and $\alpha_{1,m}$, we let $G_{1,m,n}(\bar{s}_{1,m,n}) \triangleq constant - constant\{\log_2(1 + \alpha_{1,m}\, constant)\}$. We then plot the function $G_{1,m,n}(\bar{s}_{1,m,n})$ by varying the constant. As shown in figure A, $G_{1,m,n}(\bar{s}_{1,m,n})$ is a decreasing function of $\alpha_{1,m}$.

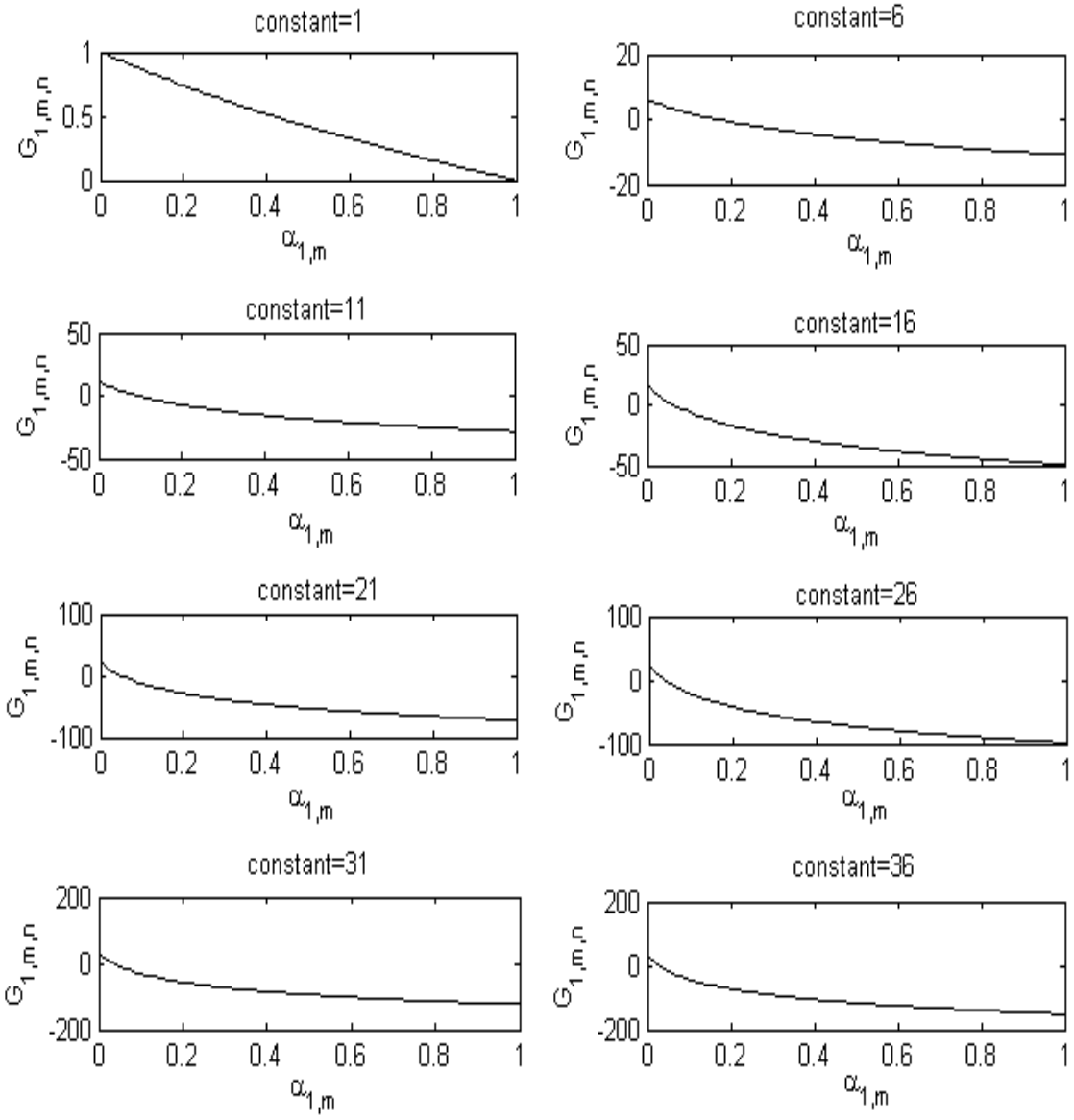


Figure A $G_{1,m,n}(\bar{s}_{1,m,n})$ versus $\alpha_{1,m}$.

APPENDIX B

Applying the natural logarithm on both sides of equation (3.4), we obtain

$$\ln \alpha_{k,n} = \ln\left(\frac{1.5}{-\ln(5BER_{k,n})}\right)$$

$$= \ln 1.5 - \ln\left(-\ln(5BER_{k,n})\right) \quad \text{(A.1)}$$

We differentiate (A.1) with respect to $BER_{k,n}$ to get

$$\left(\frac{1}{\alpha_{k,n}}\right)\left(\frac{\partial \alpha_{k,n}}{\partial BER_{k,n}}\right) = -\left(\frac{1}{-\ln(5BER_{k,n})}\right)\left(\frac{1}{BER_{k,n}}\right) \quad \text{(A.2)}$$

After some algebraic operations, (A.2) can be rewritten as

$$\left(\frac{\partial \alpha_{k,n}}{\partial BER_{k,n}}\right) = \left(\frac{1.5}{BER_{k,n} \times \ln^2(5BER_{k,n})}\right) \quad \text{(A.3)}$$

As the gradient $\left(\partial \alpha_{k,n}/\partial BER_{k,n}\right)$ is always positive for the feasible BER range is $0 < BER_{k,n} < 0.02$, it is clear that $\alpha_{k,n}$ is an increasing function of $BER_{k,n}$.

REFERENCES

3GPP TS-36.942 v.8.1.0. (2008) *Universal terrestrial radio access (e-utra); radio frequency (rf) system scenarios*.

3GPP TS 23.203 v. 8.8.0. (2009) *Policy and Charging Control Architecture*.

Ahmad, Z., Worrall, S., & Kondoz, A. (2008). Unequal Power Allocation for Scalable Video Transmission over Wimax. *2008 Ieee International Conference on Multimedia and Expo, Vols 1-4*, 517-520. doi: Doi 10.1109/Icme.2008.4607485

Al-Suhail, G. A., Wakamiya, N., & Fyath, R. S. (2006). Error-resilience of TCP-friendly video transmission over wireless channel. *2006 9th International Conference on Control, Automation, Robotics and Vision, Vols 1- 5*, 372-377.

Alasti, M., Neekzad, B., Hui, J., & Vannithamby, R. (2010). Quality of Service in WiMAX and LTE Networks. *Ieee Communications Magazine, 48*(5), 104-111. doi: Doi 10.1109/Mcom.2010.5458370

Andrews, M., Borst, S. C., Dominique, F., Jelenkovic, P. R., Kumaran, K., Ramakrishnan, K. G., & Whiting, P. A. (1998). Dynamic bandwidth allocation algorithms for high-speed data wireless networks. *Bell Labs Technical Journal, 3*(3), 30-49. doi: Doi 10.1002/Bltj.2114

Astrom, B., & Edlund, P. (2011). United States Patent No. 20110067081 A1.

Atzori, L. (2003). Transmission of JPEG2000 images over wireless channels with unequal power distribution. *Ieee Transactions on Consumer Electronics, 49*(4), 883-888. doi: Doi 10.1109/Tce.2003.1261169

Ban, D. S., Kim, C. H., & Lee, Y. H. (2008). Adaptive Grouping MMSE Channel Estimation in OFDM-Based Wireless Systems. *2008 14th Asia-Pacific Conference on Communications, (Apcc), Vols 1 and 2*, 638-642.

Banister, B. A., Belzer, B., & Fischer, T. R. (2002). Robust image transmission using JPEG2000 and turbo-codes. *Ieee Signal Processing Letters, 9*(4), 117-119. doi: Pii S1070-9908(02)05035-6

Doi 10.1109/97.1001646

Begen, A. C., & Altunbasak, Y. (2007). An adaptive media-aware retransmission timeout estimation method for low-delay packet video. *Ieee Transactions on Multimedia, 9*(2), 332-347. doi: Doi 10.1109/Tmm.2006.886282

Bertsekas. (1995). *Nonlinear Programming*: Athena Scientific.

Biglieri, E. (2005). *Coding for Wireless Channels*: Springer.

Bohge, M., Gross, J., & Wolisz, A. (2005). The potential of dynamic power and sub-carrier assignments in multi-user OFDM-FDMA cells. *GLOBECOM '05: IEEE Global Telecommunications Conference, Vols 1-6*, 2932-2936.

Caire, G., Taricco, G., & Biglieri, E. (1998). Bit-interleaved coded modulation. *Ieee Transactions on Information Theory, 44*(3), 927-946. doi: Doi 10.1109/18.669123

Cao, L. (2007). On the unequal error protection for progressive image transmission. *Ieee Transactions on Image Processing, 16*(9), 2384-2388. doi: Doi 10.1109/Tip.2007.903262

Capozzi, F., Piro, G., Grieco, L. A., Boggia, G., & Camarda, P. (2013). Downlink Packet Scheduling in LTE Cellular Networks: Key Design Issues and a Survey. *Ieee Communications Surveys and Tutorials, 15*(2), 678-700. doi: Doi 10.1109/Surv.2012.060912.00100

Caron, F., & Coulombe, S. (2012). A Maximum Likelihood Approach to Video Error Correction Applied to H. 264 Decoding. *2012 6th International Conference on Next Generation Mobile Applications, Services and Technologies (Ngmast)*, 1-6. doi: Doi 10.1109/Ngmast.2012.34

Casu, F., Cabrera, J., Jaureguizar, F., & Garcia, N. (2011). Inter-Packet Symbol Approach To Reed-Solomon FEC Codes For RTP-Multimedia Stream Protection. *2011 Ieee Symposium on Computers and Communications (Iscc)*.

Chakareski, J., & Frossard, P. (2006). Rate-distortion optimized distributed packet scheduling of multiple video streams over shared communication resources. *Ieee Transactions on Multimedia, 8*(2), 207-218. doi: Doi 10.1109/Tmm.2005.864284

Chakareski, J., & Girod, B. (2004). Computing rate-distortion optimized policies for streaming media with rich acknowledgments. *Dcc 2004: Data Compression Conference, Proceedings*, 202-211.

Chan, A., Zeng, K., Mohapatra, P., Lee, S. J., & Banerjee, S. (2010). Metrics for Evaluating Video Streaming Quality in Lossy IEEE 802.11 Wireless Networks. *2010 Proceedings Ieee Infocom*.

Chan, Y. S., Cosman, P. C., & Milstein, L. B. (2006). A cross-layer diversity technique for multicarrier OFDM multimedia networks. *Ieee Transactions on Image Processing, 15*(4), 833-847. doi: Doi 10.1109/Tip.2005.863967

Chan, Y. S., & Modestino, J. W. (2003). A joint source coding-power control approach for video transmission over CDMA networks. *Ieee Journal on Selected Areas in Communications, 21*(10), 1516-1525.

Chang, P. R. (1997). Spread spectrum CDMA systems for subband image transmission. *Ieee Transactions on Vehicular Technology, 46*(1), 80-95.

Chang, S.-H. (2011). *Joint optimization of physical and application layers for wireless multimedia communications.* (Ph.D.), UC San Diego, ProQuest, UMI Dissertation Publishing.

Chang, Z., Ristaniemi, T., & Niu, Z. S. (2014). Radio Resource Allocation for Collaborative OFDMA Relay Networks with Imperfect Channel State

Information. *Ieee Transactions on Wireless Communications, 13*(5), 2824-2835. doi: Doi 10.1109/Twc.2014.040714.131695

Chiaraluce, F., Ciccarelli, L., Gambi, E., & Spinsante, S. (2005). On the efficiency of error concealment techniques in H.264/AVC coders. *Proceedings ELMAR-2005*, 49-52.

Cho, Y. S., Kim, J., Yang, W. Y., & Kang, C. G. (2010). *MIMO-OFDM Wireless Communications with MATLAB*: Wiley-IEEE Press.

Chou, P. A., & Miao, Z. R. (2006). Rate-distortion optimized streaming of packetized media. *Ieee Transactions on Multimedia, 8*(2), 390-404. doi: Doi 10.1109/Tmm.2005.864313

Chung, S. T., & Goldsmith, A. J. (2001). Degrees of freedom in adaptive modulation: A unified view. *Ieee Transactions on Communications, 49*(9), 1561-1571.

Cisco. Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2013–2018. Retrieved March 12, 2014, from http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-520862.html

Connie, A. T., Nasiopoulos, P., Leung, V. C. M., & Fallah, Y. P. (2008). Video packetization techniques for enhancing H.264 video transmission over 3G networks. *2008 5th Ieee Consumer Communications and Networking Conference, Vols 1-3*, 800-804. doi: DOI 10.1109/ccnc08.2007.185

Cote, G., Shirani, S., & Kossentini, F. (2000). Optimal mode selection and synchronization for robust video communications over error-prone networks. *Ieee Journal on Selected Areas in Communications, 18*(6), 952-965. doi: Doi 10.1109/49.848249

Cover, T. M., & Thomas, J. A. (1991). *Elements of Information Theory*: John Wiley & Sons, Inc.

Demirtas, A. M., Reibman, A. R., & Jafarkhani, H. (2011). Performance of H.264 with Isolated Bit Error: Packet Decode or Discard? *2011 18th Ieee International Conference on Image Processing (Icip)*, 949-952.

Digital Modulation in Communications Systems –An Introduction. (1997).

Eisenberg, Y., Luna, C. E., Pappas, T. N., Berry, R., & Katsaggelos, A. K. (2002). Joint source coding and transmission power management for energy efficient wireless video communications. *Ieee Transactions on Circuits and Systems for Video Technology, 12*(6), 411-424. doi: Doi 10.1109/Tcsvt.2002.800309

Farrugia, R. A., & Debono, C. J. (2010). A Hybrid Error Control and Artifact Detection Mechanism for Robust Decoding of H.264/AVC Video Sequences. *Ieee Transactions on Circuits and Systems for Video Technology, 20*(5), 756-762. doi: Doi 10.1109/Tcsvt.2010.2045808

Farrugia, R. A., & Debono, C. J. (2011). Robust decoder-based error control strategy for recovery of H. 264/AVC video content. *Iet Communications, 5*(13), 1928-1938. doi: DOI 10.1049/iet-com.2010.0717

Goldsmith, A. (2005). *Wireless Communications*: Cambridge University Press.

H.264 JM 18.4 Reference Software (Version 18.4). Retrieved from http://iphome.hhi.de/suehring/tml/download/

Hamzaoui, R., Stankovic, V., & Xiong, Z. X. (2005). Optimized error protection of scalable image bit streams. *Ieee Signal Processing Magazine, 22*(6), 91-107. doi: Doi 10.1109/Msp.2005.1550192

Han, L., Kang, S. S., Kim, H., & In, H. P. (2013). Adaptive Retransmission Scheme for Video Streaming over Content-Centric Wireless Networks. *Ieee Communications Letters, 17*(6), 1292-1295. doi: Doi 10.1109/Lcomm.2013.043013.130326

Harri, H., & Antti, T. (2012). *LTE Advanced: 3GPP Solution for IMT-Advanced*: John Wiley & Sons Ltd.

He, K., & Cauwenberghs, G. (1999). *An area-efficient analog VLSI architecture for state-parallel Viterbi decoding.* Paper presented at the IEEE International Symposium on Circuits and Systems, Orlando.

He, Z. H., & Xiong, H. K. (2006). Transmission distortion analysis for real-time video encoding and streaming over wireless networks. *Ieee Transactions on Circuits and Systems for Video Technology, 16*(9), 1051-1062. doi: Doi 10.1109/Tcsvt.2006.881198

Hilt, B., Sarni, M., & Lorenz, P. (2009). Enhancement of channel switching scenario and IGMPv3 Protocol Implementation

Hu, S. D., Wang, H. L., Kwong, S., Zhao, T. S., & Kuo, C. C. J. (2011). Rate Control Optimization for Temporal-Layer Scalable Video Coding. *Ieee Transactions on Circuits and Systems for Video Technology, 21*(8), 1152-1162. doi: Doi 10.1109/Tcsvt.2011.2138810

Huang, J. W., Li, Z., Chiang, M., & Katsaggelos, A. K. (2008). Joint source adaptation and resource allocation for multi-user wireless video streaming. *Ieee Transactions on Circuits and Systems for Video Technology, 18*(5), 582-595. doi: Doi 10.1109/Tcsvt.2008.919109

Huang, J. W., Subramanian, V. G., Agrawal, R., & Berry, R. A. (2009). Downlink Scheduling and Resource Allocation for OFDM Systems. *Ieee Transactions on Wireless Communications, 8*(1), 288-296. doi: Doi 10.1109/T-Wc.2009.071266

ISO/IEC 14492-2 (MPEG-4 Visual). (2004) *Coding of audio-visual objects-Part 2: Visual*.

ISO/IEC 14496-10/2005. (2007) *"Advanced video coding for generic audio-visual services", ITU-T Recommendation H.264 Amendment 3, ISO/IEC 14496-10/2005: Amd 3 - Scalable extension of H.264 (SVC)*.

ISO/IEC 15444. (2004) *Information technology -- JPEG 2000 image coding system: Core coding system*.

ISO/IEC 15444-1. (2000) *JTC1/SC29/WG1: "15444-1: Information technology—JPEG 2000 image coding system—Part 1: Core coding system,"*.

ITU. Key ICT indicators for developed and developing countries and the world (totals and penetration rates) 2005-2013. Retrieved December 21, 2013, from www.itu.int/en/ITU-D/Statistics/Documents/statistics/2013/ITU_Key_2005-2013_ICT_data.xls

Jang, J. H., & Lee, K. B. (2003). Transmit power adaptation for multiuser OFDM systems. *Ieee Journal on Selected Areas in Communications, 21*(2), 171-178. doi: Doi 10.1109/Jsac.2002.807348

JasPer software reference manual. (2002) *ISO/IEC JTC 1/SC 29/WG 1 N 2415*.

Jeong, K., & Lee, J. (2007). Low complexity channel tracking for adaptive MMSE channel estimation in OFDM. *2007 41st Annual Conference on Information Sciences and Systems, Vols 1 and 2*, 548-552. doi: Doi 10.1109/Ciss.2007.4298367

Ji, X., Huang, J. W., Chiang, M., Lafruit, G., & Catthoor, F. (2009). Scheduling and Resource Allocation for SVC Streaming over OFDM Downlink Systems. *Ieee Transactions on Circuits and Systems for Video Technology, 19*(10), 1549-1555. doi: Doi 10.1109/Tcsvt.2009.2026812

Jiang, H., Zhuang, W. H., & Shen, X. M. (2005). Cross-layer design for resource allocation in 3G wireless networks and beyond. *Ieee Communications Magazine, 43*(12), 120-126.

Kambhatla, K. K. R., Kumar, S., & Lima, C. R. (2008). Video telephony performance over cellular CDMA EVDO. *2008 Ieee International Symposium on Broadband Multimedia Systems and Broadcasting*, 171-176.

Kambhatla, K. K. R., Kumar, S., Paluri, S., & Cosman, P. C. (2012). Wireless H.264 Video Quality Enhancement Through Optimal Prioritized Packet Fragmentation. *Ieee Transactions on Multimedia, 14*(5), 1480-1495. doi: Doi 10.1109/Tmm.2012.2196508

Karray, M. K. (2010). Analytical Evaluation of QoS in the Downlink of OFDMA Wireless Cellular Networks Serving Streaming and Elastic Traffic. *Ieee Transactions on Wireless Communications, 9*(5), 1799-1807. doi: Doi 10.1109/Twc.2010.05.091501

Khalek, A. A., Caramanis, C., & Heath, R. W. (2012). A Cross-Layer Design for Perceptual Optimization Of H.264/SVC with Unequal Error Protection. *Ieee Journal on Selected Areas in Communications, 30*(7), 1157-1171. doi: Doi 10.1109/Jsac.2012.120802

Kim, H., & Han, Y. N. (2005). A proportional fair scheduling for multicarrier transmission systems. *Ieee Communications Letters, 9*(3), 210-212. doi: Doi 10.1109/Lcomm.2005.03014

Kim, I. M., & Kim, H. M. (2003). An optimum power management scheme for wireless video service in CDMA systems. *Ieee Transactions on Wireless Communications, 2*(1), 81-91. doi: Doi 10.1109/Twc.2002.806367

Kiraly, C., Abeni, L., & Lo Cigno, R. (2010). Effects of P2P Streaming on Video Quality. *2010 Ieee International Conference on Communications*.

Klaue, J., Gross, J., Karl, H., & Wolisz, A. (2003). *Semantic-aware Link Layer Scheduling ofMPEG-4 Video Streams in Wireless Systems*. Paper presented at the Applications and Services in Wireless Networks Bern, Switzerland.

Klaue, J., Rathke, B., & Wolisz, A. (2003). EvalVid - A framework for video transmission and quality evaluation. *Computer Performance Evaluation: Modelling Techniques and Tools, 2794*, 255-272.

Ko, M. G., Hong, J. E., & Suh, J. W. (2012). Error Detection Scheme for the H.264/AVC using the RD Optimized Motion Vector Constraints. *Ieee Transactions on Consumer Electronics, 58*(3), 955-962.

Kondrad, L., Bouazizi, I., Vadakital, V. K. M., Hannuksela, M. M., & Gabbouj, M. (2009). Cross-Layer Optimized Transmission of H.264/SVC streams over DVB-T2 Broadcast System. *Bmsb: 2009 Ieee International Symposium on Broadband Multimedia Systems and Broadcasting, Vols 1 and 2*, 71-87.

Kumar, S., Xu, L. Y., Mandal, M. K., & Panchanathan, S. (2006). Error resiliency schemes in H.264/AVC standard. *Journal of Visual Communication and Image Representation, 17*(2), 425-450. doi: DOI 10.1016/j.jvcir.2005.04.006

Larzon, L.-A., Degermark, M., & Pink, S. (1999). *UDP lite for real time multimedia.* Paper presented at the IEEE International Conference on Communications.

Lee, H. W., & Chong, S. (2008). Downlink Resource Allocation in Multi-Carrier Systems: Frequency-Selective vs. Equal Power Allocation. *Ieee Transactions on Wireless Communications, 7*(10), 3738-3747. doi: Doi 10.1109/T-Wc.2008.061110

Lee, K. D., & Leung, V. C. M. (2009). Utility-Based Rate-Controlled Parallel Wireless Transmission of Multimedia Streams with Multiple Importance Levels. *Ieee Transactions on Mobile Computing, 8*(1), 81-92. doi: Doi 10.1109/Tmc.2008.74

Lefloch, B., Alard, M., & Berrou, C. (1995). Coded Orthogonal Frequency-Division Multiplex. *Proceedings of the Ieee, 83*(6), 982-996. doi: Doi 10.1109/5.387096

Levine, D., Lynch, W. E., & Le-Ngoc, T. (2007). Iterative joint source-channel decoding of H.264 compressed video. *2007 Ieee International Symposium on*

Circuits and Systems, Vols 1-11, 1517-1520. doi: Doi 10.1109/Iscas.2007.378699

Li, F., & Liu, G. Z. (2009). Compressed-Domain-Based Transmission Distortion Modeling for Precoded H.264/AVC Video. *Ieee Transactions on Circuits and Systems for Video Technology, 19*(12), 1908-1914. doi: Doi 10.1109/Tcsvt.2009.2031457

Li, F., Ren, P. Y., & Du, Q. H. (2012). Joint Packet Scheduling and Subcarrier Assignment for Video Communications Over Downlink OFDMA Systems. *Ieee Transactions on Vehicular Technology, 61*(6), 2753-2767. doi: Doi 10.1109/Tvt.2012.2195511

Li, G., & Liu, H. (2003). *Dynamic resource allocation with finite buffer constraint in broadband OFDMA networks.* Paper presented at the IEEE Wireless Communications and Networking, New Orleans, USA.

Li, M. D., Chen, Z. Z., & Tan, Y. P. (2011). Joint Power Allocation and Bit Loading for Enhanced Svc Video Downlink Transmissions over Sdma/Ofdma Networks. *2011 Ieee International Conference on Multimedia and Expo (Icme)*.

Li, P., Chang, Y. L., Feng, N. N., & Yang, F. Z. (2011). A Cross-Layer Algorithm of Packet Scheduling and Resource Allocation for Multi-User Wireless Video Transmission. *Ieee Transactions on Consumer Electronics, 57*(3), 1128-1134.

Liang, X., & Zhu, J. (2003, 6-9 Oct. 2003). *An adaptive subcarrier allocation algorithm for multiuser OFDM system.* Paper presented at the Vehicular Technology Conference, 2003. VTC 2003-Fall. 2003 IEEE 58th.

Lim, C., Yoo, T., Clerckx, B., Lee, B., & Shim, B. (2013). Recent Trend of Multiuser MIMO in LTE-Advanced. *Ieee Communications Magazine, 51*(3), 127-135.

Liu, Y., Ma, Q., & Zhang, H. L. (2009). Power Allocation and Adaptive Modulation for OFDM Systems with Imperfect CSI. *2009 Ieee Vehicular Technology Conference, Vols 1-5*, 1346-1349.

Liu, Y. P., Zhang, S. Y., Xu, S. C., Zhang, Y., & Liu, Y. P. (2009). Research on H.264/SVC Compressed Video Communication in 3G. *Iccsse 2009: Proceedings of 2009 4th International Conference on Computer Science & Education*, 327-332.

Lu, X. A., Wang, Y., Erkip, E., & Goodman, D. J. (2007). Total power minimization for multiuser video communications over CDMA networks. *Ieee Transactions on Circuits and Systems for Video Technology, 17*(6), 674-685. doi: Doi 10.1109/Tcsvt.2007.896627

Luo, H. Y., Ci, S., Wu, D. L., Wu, J. J., & Tang, H. (2010). Quality-Driven Cross-Layer Optimized Video Delivery over LTE. *Ieee Communications Magazine, 48*(2), 102-109.

Maani, E., & Katsaggelos, A. K. (2010). Unequal Error Protection for Robust Streaming of Scalable Video Over Packet Lossy Networks. *Ieee Transactions on Circuits and Systems for Video Technology, 20*(3), 407-416. doi: Doi 10.1109/Tcsvt.2009.2035846

Man, H., Docef, A., & Kossentini, F. (2005). Performance analysis of the JPEG 2000 image coding standard. *Multimedia Tools and Applications, 26*(1), 27-57. doi: DOI 10.1007/s11042-005-6848-5

Moccagatta, I., Soudagar, S., Liang, J., & Chen, H. (2000). Error-resilient coding in JPEG-2000 and MPEG-4. *Ieee Journal on Selected Areas in Communications, 18*(6), 899-914. doi: Doi 10.1109/49.848245

Mokari, N., Javan, M. R., & Navaie, K. (2010). Cross-Layer Resource Allocation in OFDMA Systems for Heterogeneous Traffic With Imperfect CSI. *Ieee*

Transactions on Vehicular Technology, 59(2), 1011-1017. doi: Doi 10.1109/Tvt.2009.2035131

Moon, H., & Cox, D. C. (2009). Efficient Power Allocation for Coded OFDM Systems. *Ieee Transactions on Communications, 57*(4), 943-947. doi: Doi 10.1109/Tcomm.2009.04.060304

Namburi, P., Sarac, K., & Almeroth, K. (2006). Practical utilities for monitoring multicast service availability. *Computer Communications, 29*(10), 1675-1686. doi: DOI 10.1016/j.comcom.2005.07.012

Nguyen, Q. N., Lynch, W. E., & Tho, L. N. (2010). Iterative Joint Source-Channel Decoding for H.264 Video Transmission Using Virtual Checking Method at Source Decoder. *2010 23rd Canadian Conference on Electrical and Computer Engineering (Ccece)*.

Pahalawatta, P. (2007). *Downlink packet scheduling and resource allocation for multiuser video transmission over wireless networks* (Ph.D.), NORTHWESTERN UNIVERSITY, ProQuest.

Pahalawatta, P., Berry, R., Pappas, T., & Katsaggelos, A. (2007). Content-aware resource allocation and packet scheduling for video transmission over wireless networks. *Ieee Journal on Selected Areas in Communications, 25*(4), 749-759. doi: Doi 10.1109/Jsac.2007.070511

Pande, A., Ahuja, V., Sivaraj, R., Baik, E., & Mohapatra, P. (2013). Video Delivery Challenges and Opportunities in 4G Networks. *Ieee Multimedia, 20*(3), 88-94.

Park, J., Lee, H., Lee, S., & Bovik, A. C. (2009). Optimal Channel Adaptation of Scalable Video Over a Multicarrier-Based Multicell Environment. *Ieee Transactions on Multimedia, 11*(6), 1062-1071. doi: Doi 10.1109/Tmm.2009.2026084

Pejoski, S., & Kafedziski, V. (2013). Joint Source Channel Coding Framework for Real Time H.264/AVC Video Transmission Over Wireless MIMO Channels Using Outage Probability. *2013 11th International Conference on Telecommunication in Modern Satellite, Cable and Broadcasting Services (Telsiks) Vols 1 and 2*, 221-224.

Perrine, C., Chatelllier, C., Wang, S., & Olivier, C. (2008). A joint source channel coding strategy for video transmission. *2008 3rd International Conference on Information and Communication Technologies: From Theory to Applications, Vols 1-5*, 1033-1037.

Piro, G., Grieco, L. A., Boggia, G., Fortuna, R., & Camarda, P. (2011). Two-Level Downlink Scheduling for Real-Time Multimedia Services in LTE Networks. *Ieee Transactions on Multimedia, 13*(5), 1052-1065. doi: Doi 10.1109/Tmm.2011.2152381

Proakis, J. (1995). *Digital Communications*: McGraw-Hill.

Qiu, X. X., & Chawla, K. (1999). On the performance of adaptive modulation in cellular systems. *Ieee Transactions on Communications, 47*(6), 884-895.

Rappaport, T. S. (2002). *Wireless Communications: Principles and Practice*: Prentice Hall.

Rhee, W., & Cioffi, J. M. (2000). Increase in Capacity of Multiuser OFDM System using Dynamic Subchannel Allocation. *Proceedings of theIEEE Vehicular Technology Conference*, 1085-1089.

Richardson, I. E. (2010). *The H.264 Advanced Video Compression Standard*: Wiley.

Rohling, H., & Gruneid, R. (1997). *Performance comparison of different multiple access schemes for the downlink of an OFDM communication system.* Paper presented at the IEEE Vehicular Technology Conference.

Sabir, M. F., Bovik, A. C., & Heath, R. W. (2010). Unequal Power Allocation for JPEG Transmission Over MIMO Systems. *Ieee Transactions on Image Processing, 19*(2), 410-421. doi: Doi 10.1109/Tip.2009.2032346

Sabir, M. F., Heath, R. W., & Bovik, A. C. (2009). Joint Source-Channel Distortion Modeling for MPEG-4 Video. *Ieee Transactions on Image Processing, 18*(1), 90-105. doi: Doi 10.1109/Tip.2008.2005819

Sadiq, B., Madan, R., & Sampath, A. (2009). Downlink Scheduling for Multiclass Traffic in LTE. *Eurasip Journal on Wireless Communications and Networking*. doi: Artn 510617

Sadr, S., Anpalagan, A., & Raahemifar, K. (2009). Radio Resource Allocation Algorithms for the Downlink of Multiuser OFDM Communication Systems. *Ieee Communications Surveys and Tutorials, 11*(3), 92-106. doi: Doi 10.1109/Surv.2009.090307

Schafhuber, D., & Matz, G. (2005). MMSE and adaptive prediction of time-varying channels for OFDM systems. *Ieee Transactions on Wireless Communications, 4*(2), 593-602. doi: Doi 10.1109/Twc.2004.843055

Schwarz, H., Marpe, D., & Wiegand, T. (2007). Overview of the Scalable Video Coding extension of the H.264/AVC standard. *Ieee Transactions on Circuits and Systems for Video Technology, 17*(9), 1103-1120. doi: Doi 10.1109/Tcsvt.2007.905532

Seong, K., Mohseni, M., & Cioffi, J. M. (2006). Optimal resource allocation for OFDMA downlink systems. *2006 IEEE International Symposium on Information Theory, Vols 1-6, Proceedings*, 1394-1398. doi: Doi 10.1109/Isit.2006.262075

Shakkottai, S., & Stolyar, A. L. (2002). Scheduling for multiple flows sharing a time-varying channel: the exponential rule. *Analytic Methods Applied Probability, 207*, 185-202.

Shan, W. (2009). A Joint Source Channel Coding Applied to Video for Sequences of Images. *Proceedings of 2009 International Conference on Image Analysis and Signal Processing*, 204-206.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal, 27*, 379.

Shen, Z. K., Andrews, J. G., & Evans, B. L. (2005). Adaptive resource allocation in multiuser OFDM systems with proportional rate constraints. *Ieee Transactions on Wireless Communications, 4*(6), 2726-2737. doi: Doi 10.1109/Twc.2005.858010

Smithson, G. (1998). *Introduction to digital modulation schemes.* Paper presented at the IEE Colloquium on The Design of Digital Cellular Handsets (Ref. No. 1998/240), London.

Song, G. C., & Li, Y. (2005). Cross-layer optimization for OFDM wireless networks - Part I: Theoretical framework. *Ieee Transactions on Wireless Communications, 4*(2), 614-624. doi: Doi 10.1109/Twc.2004.843065

Song, G. C., Li, Y., & Cimini, L. J. (2009). Joint Channel- and Queue-Aware Scheduling for Multiuser Diversity in Wireless OFDMA Networks. *Ieee Transactions on Communications, 57*(7), 2109-2121. doi: Doi 10.1109/Tcomm.2009.07.070394

Song, G. C., & Li, Y. G. (2005). Cross-layer optimization for OFDM wireless networks - Part II: Algorithm development. *Ieee Transactions on Wireless Communications, 4*(2), 625-634. doi: Doi 10.1109/Twc.2004.843067

Su, G. M., Han, Z., Wu, M., & Liu, K. J. R. (2006). A scalable multiuser framework for video over OFDM networks: Fairness and efficiency. *Ieee Transactions on Circuits and Systems for Video Technology, 16*(10), 1217-1231. doi: Doi 10.1109/Tcsvt.2006.883513

Subramanian, V. G., Berry, R. A., & Agrawal, R. (2010). Joint Scheduling and Resource Allocation in CDMA Systems. *Ieee Transactions on Information Theory, 56*(5), 2416-2432. doi: Doi 10.1109/Tit.2010.2040860

Sullivan, G. J., & Wiegand, T. (2005). Video compression - From concepts to the H.264/AVC standard. *Proceedings of the Ieee, 93*(1), 18-31. doi: Doi 10.1109/Jproc.2004.839617

Sun, H. F., Vetro, A., & Xin, J. (2007). An overview of scalable video streaming. *Wireless Communications & Mobile Computing, 7*(2), 159-172. doi: Doi 10.1002/Wcm.471

Sun, Y., & Xiong, Z. X. (2006). Progressive image transmission over space-time coded OFDM-based MIMO systems with adaptive modulation. *Ieee Transactions on Mobile Computing, 5*(8), 1016-1028.

Superiori, L., Nemethova, O., & Rupp, M. (2006). *Performance of a H.264/AVC error detection algorithm based on syntax analysis.* Paper presented at the International Conference on Advances in Mobile Computing & Multimedia.

Tan, S. S., Rim, M. J., Cosman, P. C., & Milstein, L. B. (2008). Adaptive Modulation for OFDM-based Multiple Description Progressive Image Transmission. *Globecom 2008 - 2008 Ieee Global Telecommunications Conference*. doi: Doi 10.1109/Glocom.2008.Ecp.620

Tan, Y. H., Lee, W. S., Tham, J. Y., & Rahardja, S. (2009). Joint Source Coding-Transmission Power Control for H.264 Video Delivery. *Tencon 2009 - 2009 Ieee Region 10 Conference, Vols 1-4*, 1387-1392.

Taubman, D. (2000). High performance scalable image compression with EBCOT. *Ieee Transactions on Image Processing, 9*(7), 1158-1170. doi: Doi 10.1109/83.847830

Taubman, D. S., & Marcellin, M. W. (2002). *JPEG2000 Image Compression Fundamentals, Standards and Practice*: Kluwer Academic Publishers.

Thomos, N., Argyropoulos, S., Boulgouris, N. V., & Strintzis, M. G. (2005). *Error-resilient transmission of H.264/AVC streams using flexible macroblock ordering*. Paper presented at the The 2nd European Workshop on the Integration of Knowledge, Semantics and Digital Media Technology.

Tosato, F., & Bisaglia, P. (2002). *Simplified Soft-output Demapper for Binary Interleaved COFDM with Application to HIPERLAN/2.* Paper presented at the IEEE International Conference on Communication.

Trudeau, L., Coulombe, S., & Pigeon, S. (2011). Pixel Domain Referenceless Visual Degradation Detection and Error Concealment for Mobile Video. *2011 18th Ieee International Conference on Image Processing (Icip)*.

Tupelly, R., Zhang, J., & CHong, E. (2003). *Opportunistic scheduling for streaming video in wireless networks.* Paper presented at the Information Sciences and Systems.

Ukhanova, A., Belyaev, E., Wang, L., & Forchhammer, S. (2012). Power consumption analysis of constant bit rate video transmission over 3G networks. *Computer Communications, 35*(14), 1695-1706. doi: DOI 10.1016/j.comcom.2012.05.010

van de Beek, J. J., Borjesson, P. O., Boucheret, M. L., Landstrom, D., Arenas, J. M., Odling, P., Wilson, S. K. (1999). A time and frequency synchronization scheme for multiuser OFDM. *Ieee Journal on Selected Areas in Communications, 17*(11), 1900-1914. doi: Doi 10.1109/49.806820

van der Schaar, M., Andreopoulos, Y., & Hu, Z. P. (2006). Optimized scalable video streaming over IEEE 802.11a/e HCCA wireless networks under delay constraints. *Ieee Transactions on Mobile Computing, 5*(6), 755-768. doi: Doi 10.1109/Tmc.2006.81

van der Schaar, M., & Turaga, D. S. (2007). Cross-layer packetization and retransmission strategies for delay-sensitive wireless multimedia transmission. *Ieee Transactions on Multimedia, 9*(1), 185-197. doi: Doi 10.1109/Tmm.2006.886384

Viterbi, A. J. (1967). Error Bounds for Convolutional Codes and an asymtotically optimal decoding algorithm. *IEEE Transaction on Information Theory, IT-13*, 260-269.

Wang, D. W., Toni, L., Cosman, P. C., & Milstein, L. B. (2013). Uplink Resource Management for Multiuser OFDM Video Transmission Systems: Analysis and Algorithm Design. *Ieee Transactions on Communications, 61*(5), 2060-2073. doi: Doi 10.1109/Tcomm.2013.032013.120053

Wang, Y., & Yu, S. Y. (2005). Joint source-channel decoding for H.264 coded video stream. *Ieee Transactions on Consumer Electronics, 51*(4), 1273-1276.

Wang, Y. X., Chen, F. J., & Wei, G. (2005). Adaptive subcarrier and bit allocation for multiuser OFDM system based on genetic algorithm. *2005 International Conference on Communications, Circuits and Systems, Vols 1 and 2, Proceedings*, 242-246.

Wehrle, K., Günes, M., & Gross, J. (2010). *Modeling and Tools for Network Simulation*: Springer.

Wenger, S. (2003). H.264/Avc over Ip. *Ieee Transactions on Circuits and Systems for Video Technology, 13*(7), 645-656. doi: Doi 10.1109/Tcsvt.2003.814966

Wong, C. Y., Cheng, R. S., Ben Letaief, K., & Murch, R. D. (1999). Multiuser OFDM with adaptive subcarrier, bit, and power allocation. *Ieee Journal on Selected Areas in Communications, 17*(10), 1747-1758.

Wong, I. C., & Evans, B. L. (2005). Joint channel estimation and prediction for OFDM systems. *GLOBECOM '05: IEEE Global Telecommunications Conference, Vols 1-6*, 2255-2259.

Wu, D. P., Hou, Y. T., Li, B., Zhu, W. W., Zhang, Y. Q., & Chao, H. J. (2000). An end-to-end approach for optimal mode selection in Internet video communication: Theory and application. *Ieee Journal on Selected Areas in Communications, 18*(6), 977-995. doi: Doi 10.1109/49.848251

Wu, Y., Kumar, S., Hu, F., Zhu, Y., & Matyjas, J. D. (2014). Cross-Layer Forward Error Correction Scheme Using Raptor and RCPC Codes for Prioritized Video Transmission Over Wireless Channels. *IEEE Transactions on Circuits and Systems for Video Technology, 24*(6), 1047 - 1060

Xu, Y. L., & Zhou, Y. H. (2004). H.264 video communication based refined error concealment schemes. *Ieee Transactions on Consumer Electronics, 50*(4), 1135-1141.

Yang, Z., & Wang, X. D. (2013). Scalable Video Broadcast Over Downlink MIMO-OFDM Systems. *Ieee Transactions on Circuits and Systems for Video Technology, 23*(2), 212-223. doi: Doi 10.1109/Tcsvt.2012.2203216

Yao, Y. W., & Giannakis, G. B. (2005). Rate-maximizing power allocation in OFDM based on partial channel knowledge. *Ieee Transactions on Wireless Communications, 4*(3), 1073-1083. doi: Doi 10.1109/Twc.2005.847022

Ye, Q. Y., Rong, B. Y., Chen, Y. D., Al-Shalash, M., Caramanis, C., & Andrews, J. G. (2013). User Association for Load Balancing in Heterogeneous Cellular

Networks. *Ieee Transactions on Wireless Communications, 12*(6), 2706-2716. doi: Doi 10.1109/Twc.2013.040413.120676

Ye, S. G., Blum, R. S., & Cimini, L. J. (2006). Adaptive OFDM systems with imperfect channel state information. *Ieee Transactions on Wireless Communications, 5*(11), 3255-3265. doi: Doi 10.1109/Twc.2006.05004

Yen, C. M., Chang, C. J., & Wang, L. C. (2010). A Utility-Based TMCR Scheduling Scheme for Downlink Multiuser MIMO-OFDMA Systems. *Ieee Transactions on Vehicular Technology, 59*(8), 4105-4115. doi: Doi 10.1109/Tvt.2010.2062546

Yu, W., & Lui, R. (2006). Dual methods for nonconvex spectrum optimization of multicarrier systems. *Ieee Transactions on Communications, 54*(7), 1310-1322. doi: Doi 10.1109/Tcomm.2006.877962

Zargari, F., & Fatemi, O. (2005). Novel error concealment methods in JPEG 2000. *Canadian Journal of Electrical and Computer Engineering-Revue Canadienne De Genie Electrique Et Informatique, 30*(2), 73-80. doi: Doi 10.1109/Cjece.2005.1541728

Zhai, F., Eisenberg, Y., Pappas, T. N., Berry, R., & Katsaggelos, A. K. (2005). Joint source-channel coding and power adaptation for energy efficient wireless video communications. *Signal Processing-Image Communication, 20*(4), 371-387. doi: DOI 10.1016/j.image.2005.02.002

Zhang, C. Y., Yang, H., Yu, S. Y., Yang, X. K., & Liu, H. (2007). GOP-level transmission distortion modeling for unequal importance judgement. *2007 Ieee International Conference on Multimedia and Expo, Vols 1-5*, 855-858.

Zhang, G. D. (2004). Subcarrier and bit allocation for real-time services in multiuser OFDM systems. *2004 Ieee International Conference on Communications, Vols 1-7*, 2985-2989.

Zhang, H. X., & Rangarajan, S. (2008). *Adaptive Scheduling of Streaming Video over Wireless Networks.* Paper presented at the IEEE International Conference on Multimedia and Expo.

Zhang, R., Regunathan, S. L., & Rose, K. (2000). Video coding with optimal inter/intra-mode switching for packet loss resilience. *Ieee Journal on Selected Areas in Communications, 18*(6), 966-976. doi: Doi 10.1109/49.848250

Zhang, X. J., Wu, Y. X., Fowler, S., Cen, S., & Dong, X. S. (2014). Joint Source-Channel Coding With Unequal Protection For Real-Time Scalable Video Transmission. *Intelligent Automation and Soft Computing, 20*(1), 77-90. doi: Doi 10.1080/10798587.2014.887459

Zhang, Y. F., Qin, S. Y., & He, Z. H. (2010). Fine-Granularity Transmission Distortion Modeling for Video Packet Scheduling Over Mesh Networks. *Ieee Transactions on Multimedia, 12*(1), 1-12. doi: Doi 10.1109/Tmm.2009.2036290

Zhu, H. L. (2012). Radio Resource Allocation for OFDMA Systems in High Speed Environments. *Ieee Journal on Selected Areas in Communications, 30*(4), 748-759. doi: Doi 10.1109/Jsac.2012.120509

Zrae, R. A., Hassan, M., & El-Tarhuni, M. (2010). *An adaptive modulation scheme for image transmission over wireless channels.* Paper presented at the IEEE International Symposium on Signal Processing and Information Technology.

LIST OF PUBLICATIONS

JOURNAL

1. **Mau-Luen Tham**, Chee-Onn Chow, Masahiro Iwahashi, and Hiroshi Ishii (2014). BER-driven resource allocation for scalable bitstreams over OFDMA networks. *IEEE Transactions on Vehicular Technology*, 63 (6), 2755-2768 (ISI-Cited Publication).

2. **Mau-Luen Tham**, Chee-Onn Chow, Yi-Han Xu, Khong Neng Choong, and Chen Suan Lee (2014). Seamless handover between unicast and multicast multimedia streams. *Journal of Zhejiang University-Science C (Computers & Electronics)*, (in press) (ISI-Cited Publication).

CONFERENCE

1. **Mau-Luen Tham**, Chee-Onn Chow, Keisuke Utsu, and Hiroshii Ishii. BER-driven resource allocation in OFDMA systems. In *Proc. IEEE 24th Personal Indoor and Mobile Radio Communications (PIMRC)*, London, 8-11 Sept. 2013.