Quality-Driven Cross Layer Design of Video Transmissions over MIMO Systems

Xiang Chen

A dissertation submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2015

Reading Committee:

Jenq-Neng Hwang, Chair

James A. Ritcey

Payman Arabshahi

Program Authorized to Offer Degree:
Electrical Engineering

0

University of Washington

Abstract

Quality-Driven Cross Layer Design of Video Transmissions over MIMO Systems

Xiang Chen

Chair of the Supervisory Committee:
Professor Jenq-Neng Hwang
Department of Chair

Mobile data traffic has been exponentially increasing during the past years. With the prevalence of video-enabled mobile devices such as smart phones, more than half of the total mobile data traffics are attributed to mobile video traffics. Since wireless resource is limited and video-related applications are bandwidth-consuming, efficiently transmitting videos over wireless networks while providing higher perceptual qualities at the end users become the everlasting endeavors of the service providers. The rapid increasing demands of wireless video streaming services have boosted the developments of video delivery technologies. In the application (APP) layer, more efficient video source coding technologies are designed so that videos with higher qualities can be delivered with less bandwidth consumptions. Among the existing video coding technologies, the scalable video coding (SVC) provides a capability of adapting to various needs or preferences of end users as well as to varying terminal capabilities or network conditions, which makes it an attractive candidate for wireless video streaming applications. In the medium access control (MAC) layer, resource allocation techniques, including rate adaptations and channel selections, are optimized for better wireless resource usages. In the physical (PHY) layer, advanced data transmission schemes such as multi-input multi-output (MIMO) are developed for higher spectral efficiency. To better utilize the advantages of different technologies in different layers, cross-layer design driven by optimizing the end users perceptral quality is highly required. In this dissertation, methods of quality-driven cross-layer design of video transmissions over MIMO systems are discussed.

First, a near optimal transmit power allocation scheme for delivering SVC-based videos

over MIMO systems is discussed. According to the channel state information (CSI) feedbacks, the power of each transmit antennas is adaptively optimized subject to an overall power limit. Therefore, the SVC bit streams with different priorities are transmitted with unequal error protections (UEPs) so that the quality of experience (QoE) at the end user is maximized. Both transmission errors in the PHY layer and video source coding characteristics in the APP layer are jointly considered in this scheme. A near optimal solution is achieved by decomposing the original optimization problem into several convex optimization sub-problems. Detailed algorithms with different complexities and their corresponding theoretical reasonings are provided. The near optimality of the proposed scheme, in terms of measured utilities, is shown by comparing with the exhaustive searched optimal solutions. Simulations with real H.264 SVC video traces demonstrate the effectiveness of the proposed scheme by comparing with other existing schemes in terms of well-accepted video quality assessment methods, such as the peak signal-to-noise ratio (PSNR) and the structural similarity (SSIM) index.

Second, a joint rate and power adaptation scheme for SVC-based video transmissions over MIMO systems is discussed. The ultimate goal of this scheme is to maximize the decoding quality at the receiver side. The rate adaptation includes selecting the best modulation and coding schemes (MCSs), set of spatial channels, number of SVC layers (source coding rates) and their corresponding application layer forward error correction (APP-FEC) coding rates. The power adaptation involves the proper allocation of the power to each antenna in the MIMO system. In most of the existing studies, the bit stream of each particular SVC layer is allocated to one spatial channel and the UEP is achieved by transmitting the more important SVC layers through the spatial channels with higher channel gains. However, in the proposed scheme, the bit stream of each particular SVC layer is distributed to multiple spatial channels so that additional diversity gain can be exploited by applying APP-FEC. The UEP can also be achieved with different APP-FEC coding rate on each video layer. Moreover, transmit power allocation is also effectively and jointly determined to improve the system performance. The effectiveness and favorable performance of the proposed scheme are shown by simulations with H.264 SVC traces of high definition (HD) video clips over MIMO systems.

TABLE OF CONTENTS

Page

LIST OF FIGURES

LIST OF TABLES

GLOSSARY

3GPP: 3rd generation partnership project

ACS: adaptive channel selection

AMC: adaptive modulation and coding

APP-FEC: application layer forward error correction

ARQ: automatic repeat request

AWGN: additive white Gaussian noise

BCR: block correction rate

BER: bit error rate

BLAST: Bell laboratories layered space-time

BPSK: binary phase-shift keying

BS: base station

CAGR: compound annual growth rate

CDF: cumulative distribution function

CIF: common interchange format

CSI: channel state information

DL: downlink

FEC: forward error correction

GF: Galois field

GOP: group of pictures

HD: high definition

HIPERLAN: high performance radio local access network

IEEE: institute of electrical and electronics engineers

IOT: Internet of things

IP: Internet protocol

IPTV: Internet protocol television

JSVM: joint scalable video model

KKT: Karush-Kuhn-Tucker

LTE: long term evolution

MAC: medium access control

MCS: modulation and coding scheme

MDS: maximum distance separable

MGS: medium-grain scalability

MIMO: multi-input multi-output

MISO: multi-input single-output

MSE: mean squared error

MS-SSIM: multi-scale-structural similarity index

MU-MIMO: multi-user multi-input multi-output

M-WF: modified water-filling

NALU: network abstraction layer unit

OFDM: orthogonal frequency-division multiplexing

OSI: open system interconnection

O-STBC: orthogonal space-time block code

PA: power allocation

PLR: packet loss rate

PMF: probability mass function

PSNR: peak signal-to-noise-ratio

QAM: quadrature amplitude modulation

QP: quantization parameter

QPSK: quadrature phase-shift keying

QOE: quality of experience

QOS: quality of service

RCPT: rate-compatable punctured turbo

RS: Reed-Solomon

RTCP: real time control protocol

RTP: real time protocol

SD: spatial diversity

SIMO: single-input multi-output

SISO: single-input single-output

SM: spatial multiplexing

SNR: signal-to-noise ratio

SSIM: structure similarity index

STBC: space-time block code

STC: space-time coding

SU-MIMO: single-user multi-input multi-output

SVC: scalable video coding

SVD: singular value decomposition

TCP: transmission control protocol

UEP: unequal error protection

UL: uplink

VQA: video quality assessment

WIMAX: worldwide interoperability for microwave access

WF: water-filling

ZMCSCG: zero-mean circularly symmetric complex Gaussian

ACKNOWLEDGMENTS

First of all, I wish to express my sincere appreciation to Prof. Jenq-Neng Hwang, my Ph.D. advisor, for his encouragement and invaluable advices during my Ph.D. career. Second, I would like to thank all my Ph.D. Supervisory Committee members: Prof. James A. Ritcey, Prof. Payman Arabshahi, Prof. John Zahorjan and Prof. Matthew Reynolds, for their helpful suggestions when I was struggled with my research topic. Moreover, I want to thank all my friends in University of Washington, especially my colleagues in the Information Processing Laboratory (IPL). Last but not least, I would like to sincerely appreciate my parents Zhengwei Chen, Hong Shi and my girl friend Yuandong Li for their understanding, kindness and support.

DEDICATION

to my parents

Chapter 1

INTRODUCTION

1.1 Motivations and Objectives

As predicted in [4], 72% of all consumer mobile traffic will be mobile video traffic in 2019, up from 55% in 2014. Furthermore, mobile data traffic will exponentially increase between 2014 and 2019, representing a 57 percent of compound annual growth rate (CAGR), which is about three times faster than fixed Internet protocol (IP) traffic. The rapid increasing demands of wireless multimedia applications have boosted the development of modern video delivery technologies over wireless channels [32]. Providing higher quality mobile video services becomes the everlasting endeavors of multimedia service providers [70]. However, the error-prone and band-limited nature of wireless communication environments, which causes high packet loss/error rate, delay and jitter, may lead to tremendous quality degradation of real-time streaming video services [27].

Many technologies have been developed to compensate for the effects caused by varying wireless channel quality. In the application (APP) layer, the scalable video coding (SVC) [63, 77] is employed so that videos can be encoded with several spatial, temporal and quality scalabilities (layers), including one base layer and several enhancement layers, where valid bit streams can still be formed even when parts of the encoded bit streams (higher enhancement layers) are removed. With more decoded enhancement layers, the received video quality is improved. But the video layers are dependent such that the base layer and lower enhancement layers are required to be successfully decoded in order to decode higher enhancement layers. Thus, SVC can be used to accommodate varying terminal capabilities or network conditions as well as satisfying the various needs or preferences of end users. Another widely used APP layer technology for video streaming is application layer forward error correction (APP-FEC) [18, 34, 48]. With the APP-FEC, data packets are transmitted along with additional well-designed redundant packets. By receiving

subsets of data and redundant packets, the receiver can reconstruct all of original data packets even when some packets are lost [34]. Therefore, APP-FEC can provide certain correction capability without retransmissions. This is suitable for delay-sensitive real-time video streaming applications [34,48]. In the physical (PHY) layer, multi-input multi-output (MIMO) technology is one of the effective solutions to support high quality video streaming services [9,14,15,17,54,56,60]. By taking advantage of spatial diversity (SD) and/or spatial multiplexing (SM), MIMO can significantly improve transmission reliability and/or spectral efficiency accordingly [15, 16, 88]. When beamforming and power allocation techniques are applied, MIMO systems can also provide unequal error protection (UEP) on bit streams transmitted through different spatial channels [10, 54]. Since SVC video layers have different decoding priorities due to their inter-layer dependency, MIMO systems are suitable for transmitting SVC-based videos with UEP on each video layer [47, 69]. Furthermore, adaptive modulation and coding (AMC) is supported in modern wireless broadband networks [3,64] in order to match transmission rates to time-varying channel conditions [6,30]. Therefore, AMC is commonly used in video streaming applications to provide heterogeneity of networks and user capabilities [9, 11, 13, 43]

To efficiently utilize the varying wireless channel, it is crucial to adaptively configure the system parameters with cross-layer design so that the characteristics of technologies on different open system interconnection (OSI) layers are jointly optimized. Various quality of service (QoS) measures, such as packet loss rate, delay and jitter etc., can be used as the criteria of network optimization for the general purpose designs of wireless transmission systems. For video streaming services, the ultimate goal is to improve the decoding quality at the user end. Since video service providers need to observe and react quickly on quality problems perceived by customers on the delivered videos, the concept of quality of experience (QoE) emerges, combining user perceptions and experiences with non-technical and technical parameters [24]. In the recent video transmission research, transmission and video coding parameters are adjusted based on well-designed objective video quality measures, which have high correlation with users' subjective perceptual satisfaction. For SVC-based video transmissions over MIMO systems, decoding errors in the high-priority video layers will cause propagation errors in the low-priority video layers. Therefore, directly minimiz-

ing bit error rate (BER) of the system does not necessarily ensure that more video frames with higher quality enhancement layers can be received. To optimize the decoding qualities at the end users, a cross-layer design of SVC-based video delivery scheme, which considers both the PHY layer characteristics and the APP layer video coding structures, is highly required to maximize the users' perceptral quality of the decoded videos.

1.2 Overview of Scalable Video Coding

The rapid developments of network infrastructures, storage capacity, and computing power, which are enabling an increasing number of video services, motivates the advances in video coding technology and standardization [63]. Real time protocol (RTP) and real time control protocol (RTCP) are used for real time streaming in modern video transmission systems over the Internet and mobile networks [58]. Based on RTCP reports, transmitters can have the statistical information of receiving devices varying connection qualities. And the variety of devices with heterogeneous receiving capabilities arise the needs for scalabilities of the encoded videos in spatial formats, bit rates or power. SVC is a highly attractive solution to the problems of modern video transmission systems, especially in adapting the needs or preferences of end users in internet protocol television (IPTV) services. Efficient SVC provides a number of benefits in terms of applications. With a properly configured SVC scheme, the source content has to be encoded only once and the representations with lower resolution and/or quality can be obtained by discarding subset of encoded data. For example, a client with restricted resources, such as display resolution, processing power or limited bandwidth, needs to decode only a part of the delivered bit stream. Moreover, in multicast services, terminals with different capabilities can be served by a single scalable bit stream. Another benefit of SVC is that a scalable bit stream usually contains parts with different importance in terms of decoded video quality. With unequal error protection (UEP) schemes, video transmissions can become more reliable over the transmission channels of unpredictable throughput variations and/or relatively high packet loss rates. SVC is also highly desirable for surveillance applications, in which video contents not only need to be viewed on variety devises, but also to be stored and archived. With SVC, low-quality copies of the video are kept for long-term archival and high-quality parts can be deleted after some

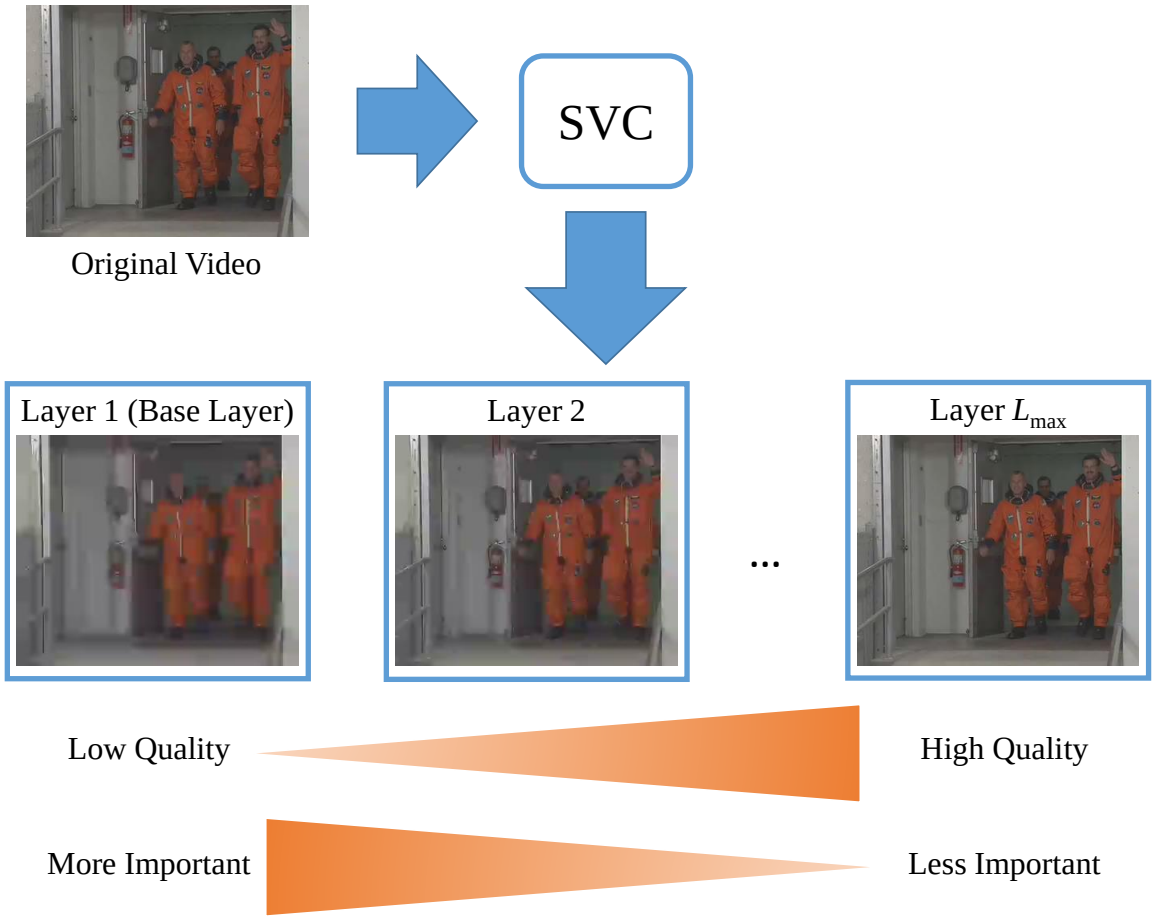

Figure 1.1: Concept of SVC-Based Video.

expiration time [78]. In the SVC extension of H.264/AVC standard, the usual modes of scalability are temporal, spatial, and quality scalabilities [77]. With spatial scalability and temporal scalability, subsets of the bit stream represent the source content with a reduced picture size (spatial resolution) and frame rate (temporal resolution) respectively. With quality scalability, the sub-stream provides the same spatial-temporal resolution, but with a lower fidelity.

Video contents can be encoded into several layers (subsets of bit streams) including one base layer and several enhancement layers as shown in Fig.1.1. The more successfully decoded enhancement layers, the better the reconstructed video qualities. Each video layer

is dependent to its corresponding "lower" layer. If $L_{\max}$ layers are encoded in total with one base layer and $L_{\max} - 1$ enhancement layers, the video decoder cannot decode the l^{th} layer if the $(l-1)^{\text{th}}$ layer is not successfully decoded. Therefore, video decoder must successfully receive the base layer bit stream in order to decode the video contents with minimum quality. Due to the inter-layer dependency characteristic of SVC-based video, in wireless video transmission, the base layer and more important enhancement layers require stronger protection compared to the less important layers.

1.3 Overview of MIMO Systems

In current wireless communication systems, MIMO is one of the key technologies for its ability to offer significant improvements in link reliability and system throughput [56]. By applying multiple antennas at transmitter and/or receiver sides, the benefits arise from the use of a new dimension space [53]. MIMO technology is also known as "space-time" wireless or "smart" antennas. Moreover, MIMO system is adopted in most of recent wireless communication standards. In the IEEE 802.16m (WiMAX) and 3rd Generation Partnership Project (3GPP) LTE-Advanced specifications, which are both considered to be the fourth generation (4G) cellular communication standards, up to 8 streams in both downlink (DL) and uplink (UL) are supported [45]. In IEEE 802.11ac (Wi-Fi) specification, which has been approved in January 2014, up to 8 spatial streams are also supported. Theoretically, the more antennas the transmitter/receiver are equipped with, the more degrees of freedom that the propagation channel can provide, which leads to the better performance in data rate or link reliability [62]. Recently, using very large antenna array at the base station (BS), also known as massive MIMO systems, attracts more and more attentions in research and developments of future wireless communication systems [29, 52].

Multiple antennas can be used to increase the SD gain to combat channel fading. Reliable reception of transmitted information can be achieved by sending signals that carry the same information through different transmit antennas and coherently combine multiple replicas of data symbols obtained at the receiver end. Extracting diversity in MIMO channels is possible with or without channel knowledge at the transmitter [55]. The SM, on the other hand, offers a linear increase in the transmission rate (or capacity) by transmitting

independent information through different transmit antennas and applying smart design of signal processing schemes at the receiver end. Given a MIMO channel, both SD and SM gains can be simultaneously obtained, but there is a fundamental tradeoff between how much of each type of gain any coding scheme can extract: higher SM gain comes at the price of sacrificing SD gain [88].

Traditional point-to-point link type of communication with multiple antennas at both transmitter and receiver sides are referred as single-user MIMO (SU-MIMO). However, in common cellular wireless systems, only one or two antennas can be equipped at the user end due to hardware limitations, while multiple antennas can be equipped on BSs. Therefore, the point-to-point link communication system can be viewed as multi-input single-output (MISO) in DL and single-input multi-output (SIMO) in UL. With one antenna at either transmitter side or receiver side, no SM gain can be exploited since SM is only possible in MIMO channels [55]. Another MIMO system mode is to deploy multiple antennas at the BS to support multiple users with one or more antennas per user terminal [55]. This type of MIMO configuration is called multi-user MIMO (MU-MIMO). In contrast to SU-MIMO, where SM gain is confined to a single user, MU-MIMO allows multiple users to be co-scheduled on the same time-frequency resources to exploit this gain among two or more users [46]. MU-MIMO system is more practical due to the scattered user distribution and lower antenna design constraints at the user end.

Numerous research in MIMO systems has been actively conducted in the past twenty years. Paujraj and Kailath [57] introduced a technique for increasing the capacity of a wireless link using multiple antennas at both ends. Bell Labs developed the BLAST architecture [25] that achieved spectral efficiencies up to 10-20 bits/s/Hz [53]. The first space-time coding (STC) architectures were introduced in [72]. And Alamouti attributed the first orthogonal space-time block code (O-STBC) for two transmit antennas to achieve full diversity with a simple signal processing scheme. Transmit beamforming techniques have been developed to increase the link signal-to-noise ratio (SNR) through focusing the energy into desired directions [23,31,87] With the emergence of IPTV system, developing transmit beamforming schemes in multicast scenarios is also a hot research area [31, 35, 66, 67].

Chapter 2

RELATED WORKS

Many studies have been conducted in rate and power allocation for video streaming applications. In [30], an opportunistic layered multicast system is proposed for scalable video multicast, where APP-FEC rates and MCS are jointly optimized for better video receiving qualities. A quality-driven resource allocation scheme for real-time video transmissions is proposed in [78], where the source coding rate and transmission data rate are jointly considered. The authors in [22] proposed a scheme for wireless video chat applications, where the system transmission parameters are adjusted based on a power-rate-distortion model.

Plenty of research has been conducted in transmitting SVC-based videos over MIMO systems, where both PHY layer structures and APP layer video coding characteristics are considered to improve the decoding quality of received videos.

Since SVC video layers are dependent with certain orders, some of the video layers are more important than the others in transmission over wireless networks. Therefore, UEP schemes, by tuning PHY layer or medium access control (MAC) layer parameters based on different optimization criteria are widely used in SVC video deliveries. In MIMO systems, UEPs can be applied by antenna selections for different video layers, choosing different transmission precoding schemes based on SD-SM tradeoffs, or assigning different transmit powers on different antennas.

In [33], the scalable video streams are transmitted over a SU-MIMO system by using STBC with equal power allocation. In this scheme, the end-to-end video distortion is minimized by optimally selecting source coding parameters including the quantization parameter (QP) and the group of picture (GoP) size. The channel coding and the physical layer parameters are also considered such as the rate-compatible punctured turbo (RCPT) channel coding rate and the symbol constellation choice for the MIMO transmission. An adaptive

channel selection (ACS) scheme is proposed in [69]. In this scheme, SU-MIMO with SM approach is considered and video bit streams with higher priorities (i.e. base layer and lower enhancement layers) are transmitted through spatial channels with relatively higher SNRs. Only partial CSIs including the orders of spatial channel gains are required at the transmitter side. As lack of full channel knowledge, equal power allocation at the transmitting antennas is assumed. A power allocation scheme, targeting on maximum throughput delivery of SVC-based videos over MIMO-SM systems, is proposed in [70]. In this scheme, the theoretical capacity-achieving water-filling (WF) algorithm is improved when discrete modulation levels are considered in a more practical scenario. However, the improvement of throughput does not directly reflect the video decoding qualities at end users. In [47], authors propose another power allocation scheme, which includes different BER targets for UEPs of transmitting different SVC video layers. Nevertheless, due to the empirical nature when setting different BER requirements on SVC video layers, this scheme may cause waste of power in different channel conditions. A resource allocation scheme is proposed in [26], where PHY layer parameters such as power and MCSs are adjusted according to video distortions measured in mean squared error (MSE). The power allocation solutions are solved by using Lagrangian method and the convexity of the optimization problem is addressed by experimental studies. A motion-compensated SVC-based video transmissions over MIMO channels is proposed in [36] , where the trade-off between error propagation and coding efficiency is exploited and the UEP is provided by FEC and MIMO mode selections. A MIMO SVC video broadcast via transmit-precoding is proposed in [79]. In this work, a precoder is computed at fixed intervals with delay and buffer constraints. And a tractable primal-dual method is developed to solve the formulated mixed-integer nonlinear optimization problem. In [54], an optimal power allocation scheme is proposed for transmitting multiple SVC-based video streams to multiple end users through MU-MIMO. An opportunistic cooperative multicast and coded cooperative multicast are proposed for scalable video communications [74]. In [44] , authors proposed a scalable resource allocation scheme for SVC video streaming over multi-user MIMO-OFDM networks.

Chapter 3

A NEAR OPTIMAL QOE-DRIVEN POWER ALLOCATION SCHEME FOR SCALABLE VIDEO TRANSMISSIONS OVER MIMO SYSTEMS

This chapter describes the proposed near optimal power allocation scheme for SVC-based video transmissions over SU-MIMO systems with SM approach. In the proposed scheme, both video coding structures in the APP layer and effects of power allocation to BER in the PHY layer are jointly considered. In practice, forward error correction (FEC) codes are often applied to improve the reliability of the communication systems [59]. Using different kind of codes with different coding rate leads to different error performances. Among plenty of error correction codes, RS code is widely used in practice and can achieve Singleton bound with its maximum distance separable (MDS) characteristic [59, 68]. Also, RS code is used in IEEE 802.16 standard [3]. In this work, three scenarios are considered: 1) system with pure quadrature amplitude modulation with constellation size M (M-QAM); 2) system with M-QAM and RS codes; 3) system with general MCSs in real applications. The third scenario referrers to the case that the standardized MCSs containing two different FECs with concatenated structure. In such cases, the close-form BER equations for each MCSs cannot be easily obtained and approximations of their BER equations [49] are considered in this work. Due to the high complexity of the original optimization problem, it is decomposed into several sub-problems which can then be solved by classical convex optimization methods. Detailed algorithms for searching the optimal solutions and its corresponding theoretical reasoning are provided. The near optimality of the proposed scheme is shown by comparing with optimal solutions through exhaustive searches. Several simulations with real H.264 SVC video traces demonstrate the effectiveness of the proposed scheme by comparing with other existing schemes in terms of well-accepted video quality assessment methods such as peak signal-to-noise ratio (PSNR) and structure similarity (SSIM) index [75].

3.1 System Overview

As shown in 3.1, a video sequence is encoded into L layers of bit streams, with one base layer and $L-1$ enhancement layers. After encoded with FECs, the bit streams are fed into a MIMO system with N_t ($N_t \geq L$) transmitter antennas and N_r receiver antennas. An adaptive channel selection (ACS) module [69, 70] is applied so that bit streams with higher priorities are transmitted through the spatial channels with higher SNRs. The power allocation module allocates appropriate power to modulated symbols with cross-layer video information and channel state information (CSI) fed back from receiver side. The data symbols are then transmitted through the wireless channel after precoding. At the receiver side, a channel estimation module sends CSIs back to the transmitter side for both power allocation and precoding processes. In this work, the accurate estimated CSIs with full channel knowledge are assumed and the CSIs are fed back without error and delay. Moreover, the channel selection sequences, modulation and FEC coding schemes are known at the receiver side through the control channel. After decoding, detection, demodulation, channel selection and FEC decoding, the received bit streams are fed into the SVC decoder for video reconstruction. For any video frame, the bit stream of video layer l is dropped if any single bit error is detected after FEC correction or the required lower layers (e.g., from base layer to the $l-1$th layer) are not successfully decoded.

3.1.1 Utility Function for SVC-Based Videos

A video bit stream is called scalable when parts of the stream can be removed in a way that the resulting sub-stream forms another valid bit stream for some target decoder [63, 77]. An SVC encoded video consists of one base layer and several enhancement layers in a hierarchical dependency structure, where the base layer and the lower enhancement layers are required in order to decode the higher enhancement layers. The decoded video quality is progressively improved when more enhancement layers are successfully decoded [86]. SVC can support all of the temporal (frame rates), spatial (picture resolutions) and quality (image fidelity) scalabilities. In this section, the proposed scheme is based on videos with quality scalability only. However, similar idea can be applied to videos with temporal and

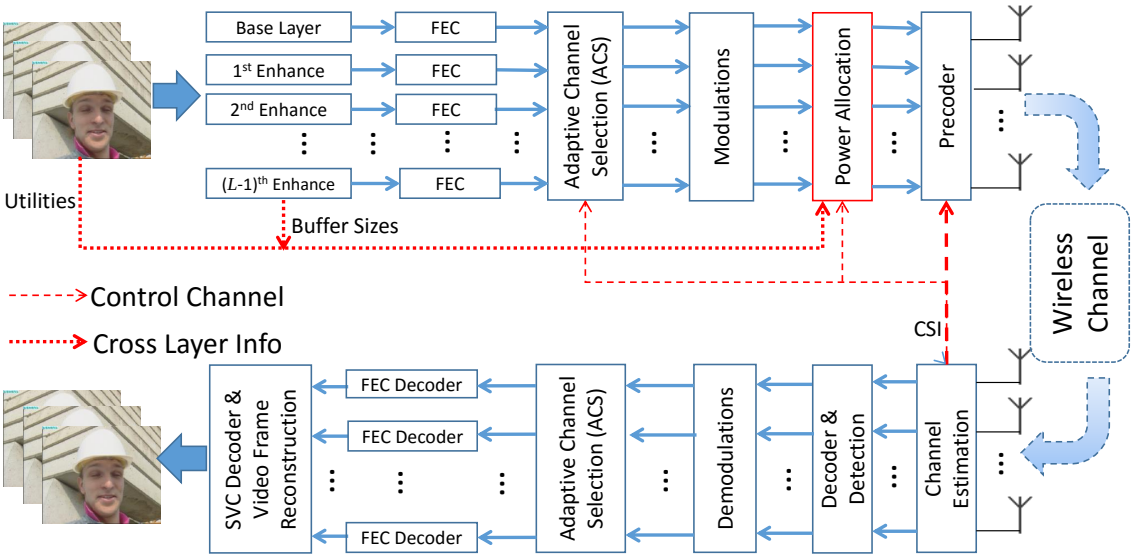

Figure 3.1: Proposed MIMO system for SVC-based video transmissions.

spatial scalabilities.

To maximize the overall QoEs at user ends, which are normally measured in utility values [30], a simple perceptual quality model for SVC-based videos with quality scalabilities [50] is adopted:

$$u_l = \begin{cases} e^{c(1-q_l/q_{\min})}, & l = 1 \\ e^{c(1-q_l/q_{\min})} - e^{c(1-q_{l-1}/q_{\min})}, & l \geq 2 \end{cases}, \tag{3.1}$$

where u_l denotes the utilities for layer l with quality scalability. c is video dependent model parameters. q_l is the quantization stepsize of the lth quality layer. $q_{\min}$ is the minimum quantization stepsize corresponding to the video layer with the highest quality. Note that the video quality model developed in [50] is specified based on videos with the common interchange format (CIF, 352×288) resolution. Therefore, videos with CIF resolution is adopted in this work. However, similar concept can be applied to videos in other resolutions with appropriate video quality models.

3.1.2 MIMO System Model

The equation of an $N_r \times N_t$ MIMO system can be described as:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}, \tag{3.2}$$

where $\mathbf{y}$ is $N_r \times 1$ received (complex) signal vector. $\mathbf{x}$ is $N_t \times 1$ complex transmitted symbol vector with $\mathbb{E}\left[\mathbf{x}\mathbf{x}^H\right] = \text{diag}\left(\mathbf{p}\right) = \text{diag}\left(p_1, p_2, ..., p_{N_t}\right)$, subject to normalized power $\mathbf{1}^T\mathbf{p} = 1$ and each element in $\mathbf{p}$ is not less than 0. $\mathbf{n}$ is an $N_r \times 1$ independent and identically distributed (*i.i.d.*) complex additive white Gaussian noise (AWGN) vector with covariance matrix $N_0\mathbf{I}_{N_r}$. $\mathbf{H}$ is $N_r \times N_t$ channel matrix in which all elements are *i.i.d.* circularly symmetric complex Gaussian (ZMCSCG) random variables with zero mean and variance 1, i.e., $\mathcal{CN}\left\{0, 1\right\}$. Therefore, the average SNR of the system is $\rho = \mathbf{1}^T\mathbf{p}/N_0 = 1/N_0$. MIMO channel matrix $\mathbf{H}$ can be decomposed by the singular value decomposition (SVD):

$$\mathbf{H} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^H, \tag{3.3}$$

where $\mathbf{U}$ and $\mathbf{V}$ are unitary matrices and $\mathbf{\Lambda}$ is a diagonal matrix specified as:

$$\mathbf{\Lambda} = \text{diag}\left(\sqrt{\lambda_1}, \sqrt{\lambda_2}, ..., \sqrt{\lambda_R}, 0, ..., 0\right), \tag{3.4}$$

where $R \leq \min\left(N_r, N_t\right)$ is the rank of channel matrix $\mathbf{H}$, and $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_R \geq 0$ are eigen values of $\mathbf{H}^H\mathbf{H}$. With accurate full channel knowledge at both transmitter and receiver side, a precoder $\mathbf{V}$ and a decoder $\mathbf{U}^H$ can be applied so that the MIMO system can be expressed as:

$$\widetilde{\mathbf{y}} = \mathbf{U}^H\mathbf{H}\mathbf{V}\mathbf{x} + \mathbf{U}^H\mathbf{n} = \mathbf{\Lambda}\mathbf{x} + \widetilde{\mathbf{n}}. \tag{3.5}$$

Since $\mathbf{U}$ is a unitary matrix, the elements in $\widetilde{\mathbf{n}} = \mathbf{U}^H\mathbf{n}$ are still *i.i.d.* complex Gaussian distributed, i.e., $\mathcal{CN}\left\{0, N_0\mathbf{I}_{N_0}\right\}$. Obviously, by using precoder and decoder, a MIMO system can be decomposed into R independent single-input single-output (SISO) channels [70]. The SNR on the lth channel can be expressed as $\text{SNR}_l = \rho\lambda_l p_l$.

3.2 Problem Formulation

The error prone nature of wireless channels will cause transmission errors in terms of bit errors or symbol errors, which further lead to video frame decoding errors and degradation

of received video qualities (i.e. system utilities). Therefore, in order to maximize QoE at the user end, the optimization problem is defined as:

$$\begin{aligned} \mathbb{Q}: \max_{\mathbf{p}} \quad & \sum_{l=1}^{L} u_l \widetilde{f}_l(\mathbf{p}) \\ \text{subject to} \quad & p_k \geq 0 \\ & \sum_{k=1}^{L} p_k = 1, \end{aligned} \tag{3.6}$$

where u_l is the utility of layer l. L is the total number of video layers. Note that $\widetilde{f}_l(\mathbf{p})$ is the frame correction rate described as:

$$\widetilde{f}_l(\mathbf{p}) = \prod_{k=1}^{l} \left(1 - \mathrm{Pe}_k(p_k)\right)^{s_k}, \tag{3.7}$$

where $\mathrm{Pe}_k(p_k)$, which is the bit or symbol error rate of the kth layer, is a decreasing function of power p_k; and s_k is the total amount of bits or symbols left in the buffer for transmitting the kth layer of a single group of pictures (GoP). In this work, a linear combination between utilities and frame correction rate is considered for simplicity. The actual utility model can vary in different scenarios such as applying APP-level error concealment techniques. However, the overall utility of a system should be an increasing function of frame correction rate. Thus, optimizing frame correction rate is a more general solution, which leads to the sub-problems discussed later.

3.2.1 Bit or Symbol Error Rates for Different MCSs

3.2.1.1 Pure M-QAM

The receiver bit error rate (BER) of M-QAM can be approximated as [68]:

$$\mathrm{Pe} \approx \frac{2\left(1 - 1/\sqrt{M}\right)}{\log_2\left(\sqrt{M}\right)} Q\left(\sqrt{\left(\frac{3\log_2\left(\sqrt{M}\right)}{M-1}\right)\frac{2E_b}{N_0}}\right), \tag{3.8}$$

where $Q(.)$ is the complementary error function and E_b/N_0 is the average bit energy to average noise power ratio. Since SNR can be calculated from E_b/N_0, i.e. $\mathrm{SNR} = \log_2(M) \times$

E_b/N_0, in the proposed MIMO-SM system, BER of the lth channel can be derived as:

$$\mathrm{Pe}_l\left(p_l\right) \approx \frac{2\left(1-1/\sqrt{M_l}\right)}{\log_2\left(\sqrt{M_l}\right)}\left(1-\Phi\left(\sqrt{\left(\frac{3}{M_l-1}\right)\rho\lambda_l p_l}\right)\right), \tag{3.9}$$

where $\Phi\left(.\right)$ is the cumulative distribution function (CDF) of the standard normal distribution.

3.2.1.2 M-QAM with RS Codes

In this paper, the proposed scheme can be also designed based on RS code. For an $(N, K, N-K+1)$ RS code over GF(2^n), which has correction capability $t=(N-K)/2$, the decoded symbol error probability is given by [59]:

$$\mathrm{Pe}_l\left(p_l\right) = \frac{1}{N}\sum_{i=t+1}^{N} i\begin{pmatrix} N \\ i \end{pmatrix}\left(P_{M_l}\left(p_l\right)\right)^i\left(1-P_{M_l}\left(p_l\right)\right)^{N-i}, \tag{3.10}$$

where $P_{M_l}\left(p_l\right)$ is the RS symbol error rate before decoding. In this chapter, the correction capability t is assumed to be an integer (i.e., $N-K$ is multiple of 2). When M-QAM modulation is applied with $\log_2\left(M\right) \leq n$, $P_{M_l}\left(p_l\right)$ can be derived as:

$$P_{M_l}\left(p_l\right) = 1-\left(1-2\left(1-1/\sqrt{M_l}\right)Q\left(\sqrt{\frac{3\rho\lambda_l p_l}{M_l-1}}\right)\right)^{\frac{2n}{\log_2\left(M_l\right)}}. \tag{3.11}$$

3.2.1.3 Approximations of MCSs in Standards

In real application standards, such as 3GPP, HiperLAN/2, IEEE 802.11a and IEEE 802.16, the BER expressions of each MCS can be approximated by [49]:

$$\mathrm{Pe}_l\left(p_l\right) = a_l e^{-b_l\rho\lambda_l p_l}, \tag{3.12}$$

where a_l and b_l are coefficients depending on the MCS scheme of layer l. Possible MCS schemes and their corresponding coefficients are listed in Table 3.1.

3.2.2 Problem Solving Strategy: Solving Sequence of Sub-problems

Due to different possible combinations of utilities and frame correction rate, a general solution is to optimize the frame correction rate directly when different numbers of SVC

Table 3.1: Possible MCSs and Corresponding Coefficients

Type(m)	**Modulation**	**Code Rate**	$\mathbf{a_m}$	$\mathbf{b_m}$
1	BPSK	1/2	1.1369	7.5556
2	QPSK	1/2	0.3351	3.2543
3	QPSK	3/4	0.2197	1.5244
4	16 QAM	9/16	0.2081	0.6250
5	16 QAM	3/4	0.1936	0.3484
6	64 QAM	3/4	0.1887	0.0871

layers are considered. Therefore, the original problem $\mathbb{Q}$ can be decomposed into L sub-problems $\mathbb{Q}_l$, where $l = 1, 2, ..., L$. More specifically, when up to the lth SVC layer is allowed to be transmitted, the corresponding frame correction rate of layer l can be optimized by solving the following sub-problem:

$$\begin{aligned} \mathbb{Q}_l : \min_{\mathbf{p}} \quad & -\sum_{k=1}^{l} s_k \log\left(1 - \mathrm{Pe}_k\left(p_k\right)\right) \\ \text{subject to} \quad & p_k \geq 0 \\ & \sum_{k=1}^{l} p_k = 1. \end{aligned} \tag{3.13}$$

Note that $p_{l+1} = p_{l+2} = ... = p_L = 0$ are implied since the layers higher than l are not allowed to be transmitted. If $\mathbf{p}_l^*$ denotes the solution of the lth sub-problem in Eq. (3.13), the optimal solution of Eq. (3.6) is found by:

$$\mathbf{p}^* = \arg\max_{\mathbf{p}_l^*} \sum_{k=1}^{L} u_k \widetilde{f}_k\left(\mathbf{p}_l^*\right). \tag{3.14}$$

Since the original problem $\mathbb{Q}$ is solved by choosing the best solution among the solutions of the corresponding sub-problems, the solution found in Eq. (3.14) is not globally optimal.

In Section 3.5, the near-optimality of the proposed scheme is demonstrated by comparing with the global optimal points obtained by exhaustive searches.

3.2.3 Convexity of Sub-problems

In this subsection, the objective function of $\mathbb{Q}_l$ in Eq. (3.13) is shown to be convex. Thus, with linear equality and inequality constraints, the sub-problem $\mathbb{Q}_l$ can be solved by classical convex optimization.

3.2.3.1 Pure M-QAM

For pure M-QAM case without any FECs, in Eq. (3.9), $\Phi(.)$ is concave and non-decreasing function for any non-negative input argument. And the argument inside $\Phi(.)$ is concave with respective to non-negative variable p_l. $\mathrm{Pe}_l(p_l)$ in Eq. (3.9) is convex due to the composition rule: $f(x) = h(g(x))$ is concave if h is concave and non-decreasing, and g is concave; and the property: $f(x)$ is convex if $-f(x)$ is concave [5]. The above composition rule can be used again to show $\log(1 - \mathrm{Pe}_l(p_l))$ is concave since $\log(.)$ is concave and non-decreasing for any non-negative input. Therefore, the objective function of $\mathbb{Q}_l$ is convex since non-negative weighted sums preserve convexity [5].

3.2.3.2 M-QAM with RS Codes

When the RS symbol error rate $P_{M_l}(p_l)$ is low enough, which is the general case in practice, the decoded symbol error rate in Eq. (3.10) can be approximated as (see Appendix A.1 for derivations):

$$\mathrm{Pe}_l(p_l) \approx \begin{cases} \dfrac{(N-1)!}{(N-t-1)} P_{M_l}(p_l)^{t+1}, & 0 \le t \le 2 \\ \displaystyle\sum_{j=3}^{t+1} \left(\frac{(N-1)!\,(j-2)}{(N-j)!\,(j-1)!} P_{M_l}(p_l)^{j} \right) + \frac{(N-1)! P_{M_l}(p_l)^{t+1}}{t!\,(N-t-2)!}, & t \ge 3 \end{cases} . \tag{3.15}$$

Define $\mathrm{Ps}_l(p_l)$ as:

$$\mathrm{Ps}_l(p_l) = 2\left(1 - 1/\sqrt{M_l}\right) Q\left(\sqrt{\frac{3\rho\lambda_l p_l}{M_l - 1}}\right), \tag{3.16}$$

which is convex by similar proof in pure M-QAM case discussed above. $P_{M_l}(p_l)$ in Eq. (3.11) can be expressed as:

$$P_{M_l}(p_l) = 1 - (1 - \mathrm{Ps}_l(p_l))^{\frac{2n}{\log_2(M_l)}} \approx \frac{2n\mathrm{Ps}_l(p_l)}{\log_2(M_l)} \tag{3.17}$$

by first order Taylor series approximation since $\mathrm{Ps}_l(p_l)$ is much smaller than 1 in practice. Therefore, by composition rule: $f(x) = h(g(x))$ is convex if h is convex and non-decreasing, and g is convex, $(P_{M_l}(p_l))^j$ is convex for $j \geq 1$. Therefore, $\mathrm{Pe}_l(p_l)$ is convex since it is non-negative sums of convex functions. Therefore the objective function of $\mathbb{Q}_l$ is conex for M-QAM with RS codes.

3.2.3.3 Approximations of MCSs in Standards

$\mathrm{Pe}_l(p_l)$ is convex with positive a_l and b_l in Eq. (3.12). Therefore, with similar proof in the pure M-QAM case, the objective function of $\mathbb{Q}_l$ is convex with the BER approximation equation for different MCSs.

3.3 Conditions of Optimal Solutions for Sub-problems

The Lagrangian of the lth sub-problem $\mathbb{Q}_l$, in Eq. (3.13), can be derived as:

$$\mathcal{L}_l(\mathbf{p}, \xi, \nu) = -\sum_{k=1}^{l} s_k \log(1 - \mathrm{Pe}_k(p_k)) - \sum_{k=1}^{l} \xi_k p_k + \nu\left(\sum_{k=1}^{l} p_k - 1\right), \tag{3.18}$$

where ξ and ν are Lagrange multipliers associated with the inequality constraints and equality constrant respectively. For each $k = 1, ..., l$, the Karush-Kuhn-Tucker (KKT) conditions can be expressed as:

1. Primal feasible: $p_k^* \geq 0$; $\sum p_k^* = 1$.

2. Dual feasible: $\xi_k^* \geq 0$

3. Complementary slackness: $\xi_k^* p_k^* = 0$.

4. Gradient of Lagrangian vanishes:

$$\frac{\partial \mathcal{L}(\mathbf{p}, \xi, \nu)}{p_k}\Big|_{p_k^*, \xi_k^*, \nu^*} = \frac{s_k \mathrm{Pe}_k'(p_k^*)}{1 - \mathrm{Pe}_k(p_k^*)} - \xi_k^* + \nu^* = 0, \tag{3.19}$$

where $\mathrm{Pe}_k'(p_k)$ is the derivative of $\mathrm{Pe}_k(p_k)$ For convex optimization problems, if any point satisfies the KKT conditions, it is primal and dual optimal with zero duality gap [5]. The above KKT conditions imply:

$$\nu^* \geq -\frac{s_k \mathrm{Pe}_k'(p_k^*)}{1-\mathrm{Pe}_k(p_k^*)}, \tag{3.20}$$

and

$$\left(\frac{s_k \mathrm{Pe}_k'(p_k^*)}{1-\mathrm{Pe}_k(p_k^*)} + \nu^*\right) p_k^* = 0. \tag{3.21}$$

Here, the case when $p_k^* > 0$ is considered since the case $p_k^* = 0$ can be solved by the $(k-1)$th sub-problem. Therefore, Eq. (3.21) implies:

$$\nu^* = \frac{1}{h_k(p_k^*)} = -\frac{s_k \mathrm{Pe}_k'(p_k^*)}{1-\mathrm{Pe}_k(p_k^*)}. \tag{3.22}$$

For pure M-QAM case,

$$\mathrm{Pe}_k'(p_k^*) \approx -\frac{\left(1-1/\sqrt{M_k}\right) e^{-\frac{3\rho\lambda_k p_k^*}{2(M_k-1)}}}{\log_2\left(\sqrt{M_k}\right)} \sqrt{\frac{3\rho\lambda_k}{2\pi(M_k-1)p_k^*}}. \tag{3.23}$$

For M-QAM with RS codes (see Appendix),

$$\mathrm{Pe}_k'(p_k^*) = \sum_{j=t}^{N-1} \frac{P_{M_k}'(p_k^*)\binom{N-1}{j}(j+1-NP_{M_k}(p_k^*))P_{M_k}(p_k^*)}{\left(1-P_{M_k}(p_k^*)\right)^{j-N+2}}, \tag{3.24}$$

where

$$P_{M_k}'(p_k^*) = \frac{2n\mathrm{Ps}_k'(p_k^*)(1-\mathrm{Ps}_k(p_k^*))^{\frac{2n}{\log_2(M_k)-1}}}{\log_2(M_k)}, \tag{3.25}$$

and

$$\mathrm{Ps}_k'(p_k^*) = -\left(1-1/\sqrt{M_k}\right) e^{-\frac{3\rho\lambda_k p_k^*}{2(M_k-1)}} \sqrt{\frac{3\rho\lambda_k}{2\pi(M_k-1)p_k^*}}. \tag{3.26}$$

For appximated BER equations of MCSs in real standards,

$$\mathrm{Pe}_k'(p_k^*) = -a_k b_k \rho\lambda_k e^{-b_k\rho\lambda_k p_k^*}. \tag{3.27}$$

Hence, the optimality conditions of the lth sub-problem $\mathbb{Q}_l$ is:

1. $h_k(p_k^*) = 1/\nu^*$ for $k = 1, 2, ..., l$; and
2. $\sum_{k=1}^{l} p_k^* = 1$; and
3. $p_k > 0$ for $k = 1, 2, ..., l$ and $p_k = 0$ for $k = l+1, l+2, ..., L$.

3.4 Proposed Algorithms

In this chapter, a simple but effective bisection search algorithm is adopted to obtain the optimal points of the sub-problems. A low complexity algorithm is also proposed, which can achieve the near optimal solutions of the sub-problems.

3.4.1 Monotonicity of Function $h(.)$

The first derivative of $h_k(p_k)$ can be derived as:

$$h'_k(p_k) = \frac{s_k \mathrm{Pe}'_k(p_k)^2 + s_k(1 - \mathrm{Pe}_k(p_k^*))\,\mathrm{Pe}''_k(p_k)}{(s_k \mathrm{Pe}'_k(p_k))^2}. \tag{3.28}$$

Since $\mathrm{Pe}_k(p_k)$ is convex as proved above, its second derivative $Pe''_k(p_k) \geq 0$. Also, $\mathrm{Pe}_k(p_k)$ is a probability of error, which is between 0 and 1, the term $(1 - \mathrm{Pe}_k(p_k))$ is non-negative. Thus, $h'_k(p_k)$ is non-negative for any power p_k (positive for the kth sub-problem) and positive s_k. Since $\mathrm{Pe}_k(p_k)$ is decreasing with respect to power p_k, its first derivative $\mathrm{Pe}'_k(p_k) < 0$. Therefore, function $h_k(p_k)$ is monotonically increasing.

3.4.2 Bisection Search Algorithm

Since $h_k(p_k)$ is monotonically increasing and $0 < p_k \leq 1$, the optimal point of the lth sub-problem is between 0 and $\min(h_k(1))$ for $k = 1, 2, ..., l$. In practice, the function $g_k(p_k) \triangleq \log(1 + h_k(p_k))$ is used in the algorithm to avoid numerical limitations without sacrificing the optimality of the lth sub-problem $\mathbb{Q}_l$. Therefore, the first optimal condition of $\mathbb{Q}_{\ll}$ becomes:

$$g_k(p_k) = \mu \triangleq \log(1 + 1/\nu^*). \tag{3.29}$$

The proposed bisection search algorithm, used to find the optimal point $\mathbf{p}_l^*$ of the lth sub-problem $\mathbb{Q}_l$, is shown in Algorithm 1. Here, ϵ is a small positive number for the stopping criterion. The first step is used to determine the upper bound of searching region, which is found by the minimum value of each $g_k(p_k)$ when the power reaches its constraint, i.e., $p_k = 1$. Step 13 is used to satisfy the power constraint without any numerical error.

Algorithm 1 (A1): Bisection Search Algorithm

1. $upper := \min\left(g_k\left(1\right)\right)$, for $k = 1, 2, ..., l$
2. $lower := 0$;
3. $p_k := 0$, for $k = 1, 2, ..., L$
4. while$\left(\left|\sum p_k - 1\right| > \epsilon\right)$
5. $\quad \mu := \left(upper + lower\right)/2$;
6. $\quad p_k := g_k^{-1}\left(\mu\right)$, for $k = 1, 2, ..., l$
7. $\quad$ if $\left(\sum_{k=1}^{l} p_k < 1\right)$
8. $\quad\quad lower := \mu$;
9. $\quad$ else
10. $\quad\quad upper := \mu$;
11. $\quad$ end if
12. end while
13. $p_k^* := p_k / \sum p_k$, for $k = 1, 2, ..., l$
14. $\mathbf{p}_l^* := \left[p_1^*, p_2^*, ..., p_L^*\right]^T$;

3.4.3 Low-Complexity Algorithm

In practice, it can be observed that for the lth sub-problem $\mathbb{Q}_l$, most of the elements p_k^* of the optimal solution $\mathbf{p}_l^*$ obtained from A1, are located at the linear region of $g_k(p_k) = \log(1 + h_k(p_k))$ for $k = 1, 2, ..., l$. This phenomenon is due to the fact that $h_k(p_k)$, defined in Eq. (3.22), is dominated by the exponential terms in Eq. (3.23), Eq. (3.24) and Eq. (3.27). for pure M-QAM, M-QAM with RS codes and MCS cases respectively. By taking advantage of this observation, $g_k(p_k)$ can be approximated as: $g_k(p_k) = A_k p_k + B_k$, where A_k and B_k are two constants to be determined. If A_k and B_k are known, the optimal constant μ^* can be directly calculated:

$$\mu^* = \frac{1 + \sum_{k=1}^{l} \frac{B_k}{A_k}}{\sum_{k=1}^{l} \frac{1}{A_k}}. \tag{3.30}$$

And each element of the optimal solution $\mathbf{p}_l^*$ can be determined as: $p_k^* = (\mu^* - B_k)/A_k$, for $k = 1, 2, ..., l$. Algorithm 2 (A2) is proposed to further reduce the complexity of A1 without sacrificing the system performance too much. Here, step 4 is used to roughly estimate the optimal solution. Step 5 and 6 are used to estimate the constants A_k and B_k, where Δ is a small positive number. Note that there is no iteration process involved in A2.

3.5 Simulation Results

In this section, the near-optimality and the effectiveness of the proposed algorithm are evaluated through plenty of simulations. Video clips "City", "Foreman" and "Waterfall" with CIF resolutions are encoded by the reference SVC codec JSVM (Joint Scalable Video Model) version 9.19. Maximum frame rates are all set as 30fps. Both GoP sizes and intra periods are set as 8 so that the frame pattern is IBBBBBBB in one GoP. There are 161 frames encoded in total so that 20 GoPs with one additional I frame are included. Three additional enhancement layers are encoded with constrained medium-grain scalability (MGS), where motion estimation and compensation are constrained at current layer [50]. The basis quantization parameters of the four layers (i.e., one base layer and three enhancement layers) are set as QP $= [48, 42, 36, 30]$ with corresponding uniform quantization stepsizes calculated by $q = 2(QP - 4)/6$ [50]. Based on Eq. (3.1), the utilities of the four quality

Algorithm 2 (A2): Low-Complexity Algorithm

1. $p_k := 0$, for $k = 1, 2, ..., L$
2. $\mu := \min\left(g_k\left(1\right)\right)$, for $k = 1, 2, ..., l$
3. $p_k := g_k^{-1}\left(\mu\right)$, for $k = 1, 2, ..., l$
4. $p_k := p_k / \sum p_k$, for $k = 1, 2, ..., l$
5. $\mu := \max\left(g_k\left(p_k\right)\right)$, for $k = 1, 2, ..., l$
6. $A_k := \left(\mu - g_k\left(g_k^{-1}\left(\mu\right) - \Delta\right)\right) / \Delta$, for $k = 1, 2, ..., l$
7. $B_k := \mu - A_k g_k^{-1}\left(\mu\right)$, for $k = 1, 2, ..., l$
8. $\mu^* := \left(1 + \sum_{k=1}^{l} B_k A_k^{-1}\right) / \sum_{k=1}^{l} A_k^{-1}$
9. $p_k^* := \left(\mu^* - B_k\right) / A_k$, for $k = 1, 2, ..., l$
10. $\mathbf{p}_l^* := \left[p_1^*, p_2^*, ..., p_L^*\right]^T$

Table 3.2: Utilities of Different Video Layers

Video	**c** [50]	$\mathbf{u_1}$	$\mathbf{u_2}$	$\mathbf{u_3}$	$\mathbf{u_4}$
City	0.13	0.4025	0.2745	0.2010	0.1219
Foreman	0.12	0.4317	0.2660	0.1892	0.1131
Waterfall	0.15	0.3499	0.2877	0.2231	0.1393

layers of "City", "Foreman" and "Crew" are listed in Table 3.2. The encoded network abstraction layer units (NALUs) areis packetized by link layer with packet size as 48 bytes [7] and then transmitted through the PHY layer. A 4×4 MIMO system is used for transmitting SVC videos with 4 layers [26, 33, 69, 70]. The channel bandwidth, modulation and channel coding rates are selected such that the source coding bitrate of each video layer are less than the corresponding throughput of each spatial channel. Table 3.3 lists the bit rate of each video layer. The bandwidths, modulation and channel coding schemes are shown in Table 3.4. The CSIs are fed back every channel coherence time, which is assumed to be 1ms [70]. At the receiver side, control messages such as video coding parameters are assumed to be correctly received. Also, perfect error detection scheme is assumed so that bit or symbol errors are correctly detected at the receiver side. The undecodable NALUs, including erroneous bits caused by channel quality degradation or unsatisfied video layer dependencies, are discarded before passing through the SVC decoder. The missed video frames are concealed by simply copying the previous successfully received frames. In the simulations, the proposed scheme is compared with traditional WF algorithm, simple equal power allocation scheme and modified WF (M-WF) proposed in [47] with BER targets set as 10^{-5} and 10^{-3} for base layer and enhancement layers respectively. Figure 3.2 illustrates the system utilities calculated by the objective function in Eq. (3.6) when transmitting the first GoP of video clip "City" with pure M-QAM modulations. The optimal curve is obtained by exhaustive searches. As shown in Fig. 3.2, the proposed algorithms are very

Table 3.3: Bit Rate (in Kbps) of Different Video Layers

Video	**Layer 1**	**Layer 2**	**Layer 3**	**Layer 4**
City	214.9	102.3	326.4	613.4
Foreman	204.3	69.2	190.6	379
Waterfall	249.3	77	265.7	566.3

Table 3.4: Bandwidth and Modulation Coding Schemes

	City (M-QAM)	**Foreman (M-QAM + RS code)**	**Waterfall (MCS)**
Bandwidth (KHz)	110	70	180
Layer 1	QPSK	16-QAM $(255, 187, 69)$	QPSK 3/4
Layer 2	QPSK	QPSK (255,127,129)	BPSK 1/2
Layer 3	16-QAM	16-QAM (255,175,81)	QPSK 3/4
Layer 4	64-QAM	64-QAM (255,231,25)	64-QAM 3/4

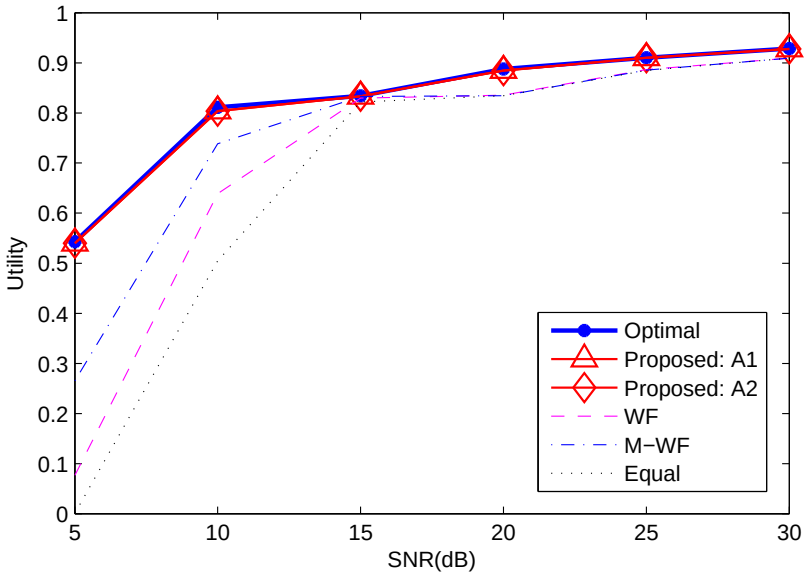

Figure 3.2: System utility of each power allocation scheme for transmitting the first GoP of video "City" with pure M-QAM modulations.

close to the optimal solutions. The performance reduction of the proposed low-complexity algorithm (A2) is negligible comparing with that of the proposed bisection search algorithm (A1). Even though the WF algorithm is optimal in terms of the PHY layer capacities, it is no longer optimal in the APP layer utilities. The M-WF scheme is better than the WF scheme at low SNR region since unequal error protection (UEP) is applied on the base layer and enhancement layers for better effect when channel qualities are not good enough. But it becomes worth at high SNR region due to the over protection of high priority layers. Similar results are plotted in Fig. 3.3 and Fig. 3.4. In Fig. 3.3, transmitting the first GoP of video clip "Foreman" with M-QAM modulations and RS codes over GF $\left(2^8\right)$ is considered. In Fig. 3.4, transmitting the first GoP of video clip "Waterfall" with different MCSs in Table 3.1 is considered. It is clear that the near optimality of the proposed scheme still holds when different MCSs are applied on different video layers.

Since the objective of the proposed scheme is to maximize the system utility, there will be no surprises that the proposed scheme can achieve higher utilities than other schemes. Next, these schemes are compared in terms of traditional video quality assessment methods

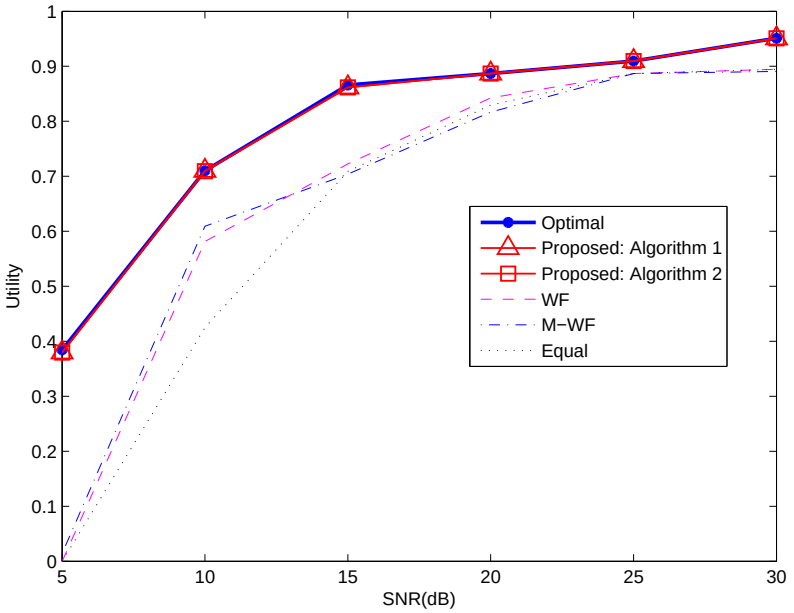

Figure 3.3: System utility of each power allocation scheme for transmitting the first GoP of video "Foreman" with M-QAM modulations and RS codes.

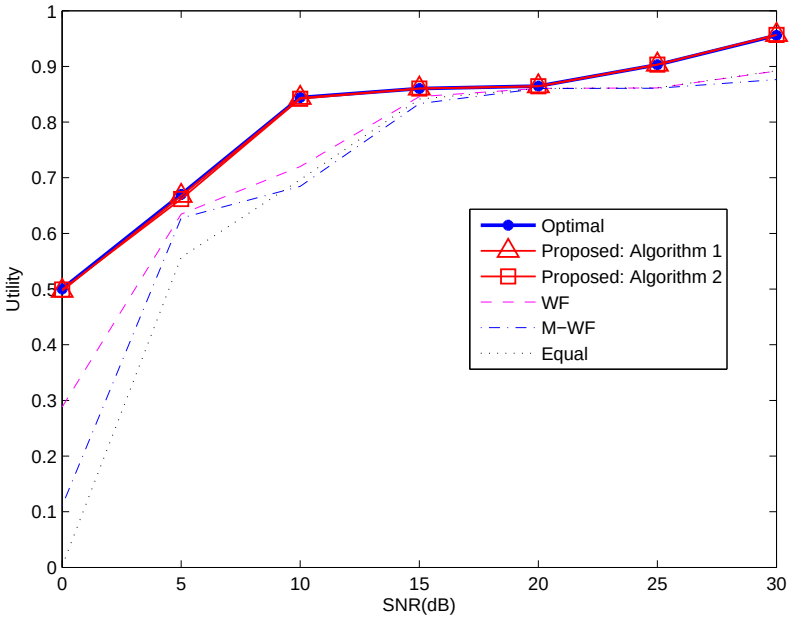

Figure 3.4: System utility of each power allocation scheme for transmitting the first GoP of video "Waterfall" with MCSs in Table 3.1.

such as PSNR and SSIM index.

The SVC layer indices of received frames of "City" video clip are plotted in Fig. 3.5. The system average SNR is set as 24dB. Clearly, more video frames with higher layers can be received by using the proposed scheme, which can improve the qualities of the reconstructed videos. The cumulative distribution function (CDF) of the power allocation results obtained from different schemes are illustrated in Fig. 3.6. One can observe that with empirical settings of BER targets, M-WF scheme allocates more power on the high priority layers (layer 1 and layer 2) while the proposed scheme allocates more power on layer 3. When the system average SNR is 24dB, channel gains of the high priority layers are normally high enough. Therefore, allocating more power on the high priority layers may cause waste of power. Figure 3.7 shows the per-frame PSNRs of the reconstructed "City" video clip. It is obvious that the proposed two algorithms outperform the other three, even though the optimization objective function is not PSNR. This is because of the fact that by applying the proposed scheme with reasonable utility functions, more video frames with higher quality layers can be received. The corresponding SSIM indices are plotted in Fig. 3.8, which still demonstrates the effectiveness of the proposed scheme. Theoretically, no matter what kind of video quality assessment methods are used, as long as they are monotonically increasing functions of video frame correction rates, the proposed algorithms can achieve higher performance than the others since more video frames with higher layers can be received. Decoded sample video frames are shown in Fig. 3.9, which clearly demonstrates the visual quality advantage of the proposed scheme. Figure 3.10 shows the average PSNR and SSIM with respect to system average SNR. The system performance degradation of the proposed low-complexity algorithm (A2) is neglegible comparing with the proposed bisection search algorithm (A1). Note that M-WF algorithm has special UEP on each video layer with empirically chosen BER targets, which would provide stronger protection on high priority video layers when channel qualities are bad. However, when the channel qualities are generally good, it would over-protect the high priority layers, which may degrade the QoE of end users since not enough higher enhancement layers can be received.

Similar simulations can be repeated on the system with M-QAM and RS codes. The

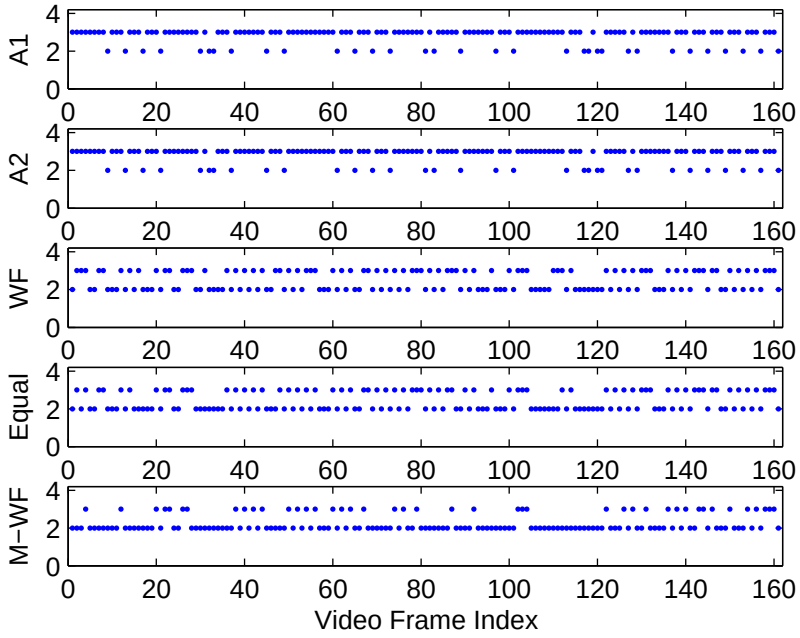

Figure 3.5: SVC quality layer indices of received frames of video "City" at system average SNR: 24dB.

SVC layer indices of received NALUs of "Foreman" video clip are plotted in Fig.3.11 where the system average SNR is set as 18dB. By using the proposed scheme, more video frames with higher quality layers can be received. The CDF of power allocation results obtained by different schemes are plotted in Fig. 3.12. Figures 3.13 and 3.14 show the per-frame PSNRs and SSIM of reconstructed "Foreman" video clip respectively. The better effectiveness of the proposed scheme is again observed comparing with that of other methods. Note that in this case, the differences of either PSNR or SSIM comparisons are smaller. This is due to the fact that the background of "Foreman" is smoother than that of "City", which causes less difference between each quality layers with different quantization stepsizes. A visual comparison of the reconstructed video frames is shown in Fig. 3.15. It is still clear that the frame is sharper by applying the proposed scheme. The average PSNR and SSIM with respect to system average SNR are plotted in Fig. 3.16. Note that in this scenario, the system performance degradation of the proposed low-complexity algorithm (A2) is more

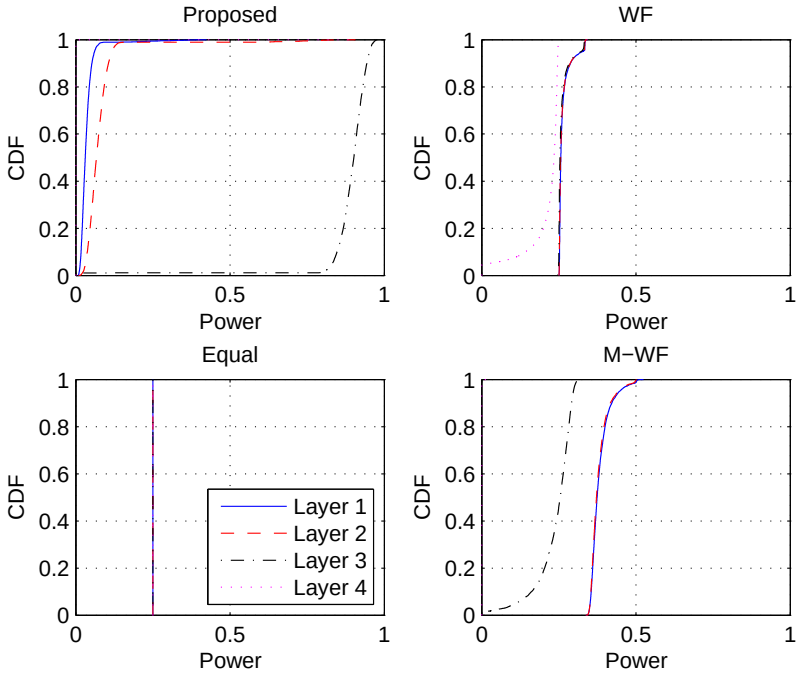

Figure 3.6: CDF of power allocation results when transmitting video "City" at system average SNR: 24dB.

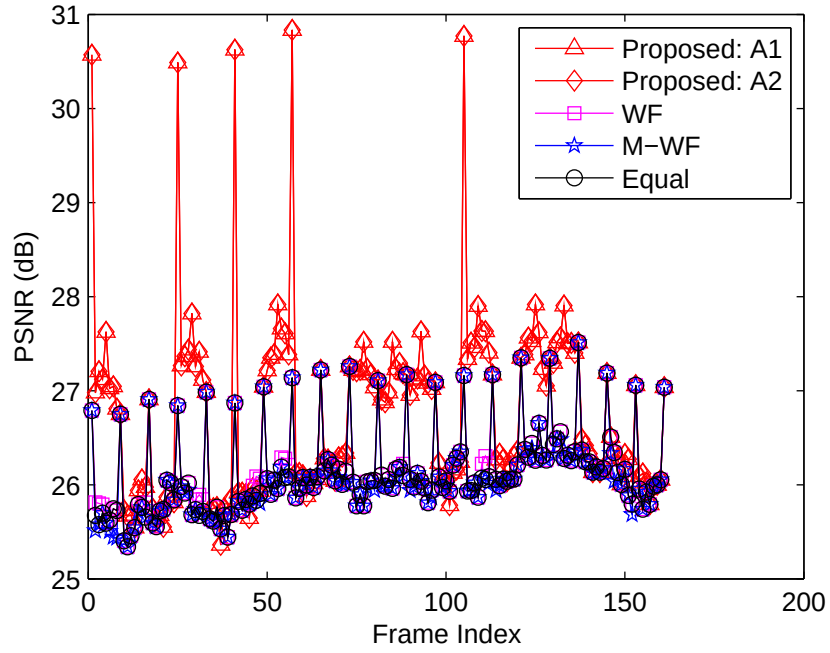

Figure 3.7: Per-frame PSNR of reconstructed video "City" at system average SNR: 24dB.

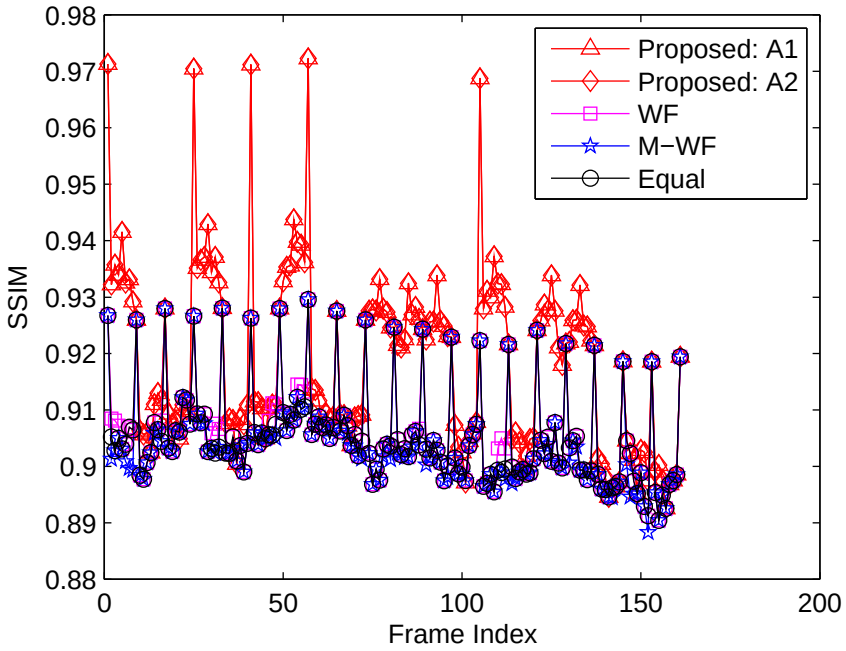

Figure 3.8: Per-frame SSIM of reconstructed video "City" at system average SNR: 24dB.

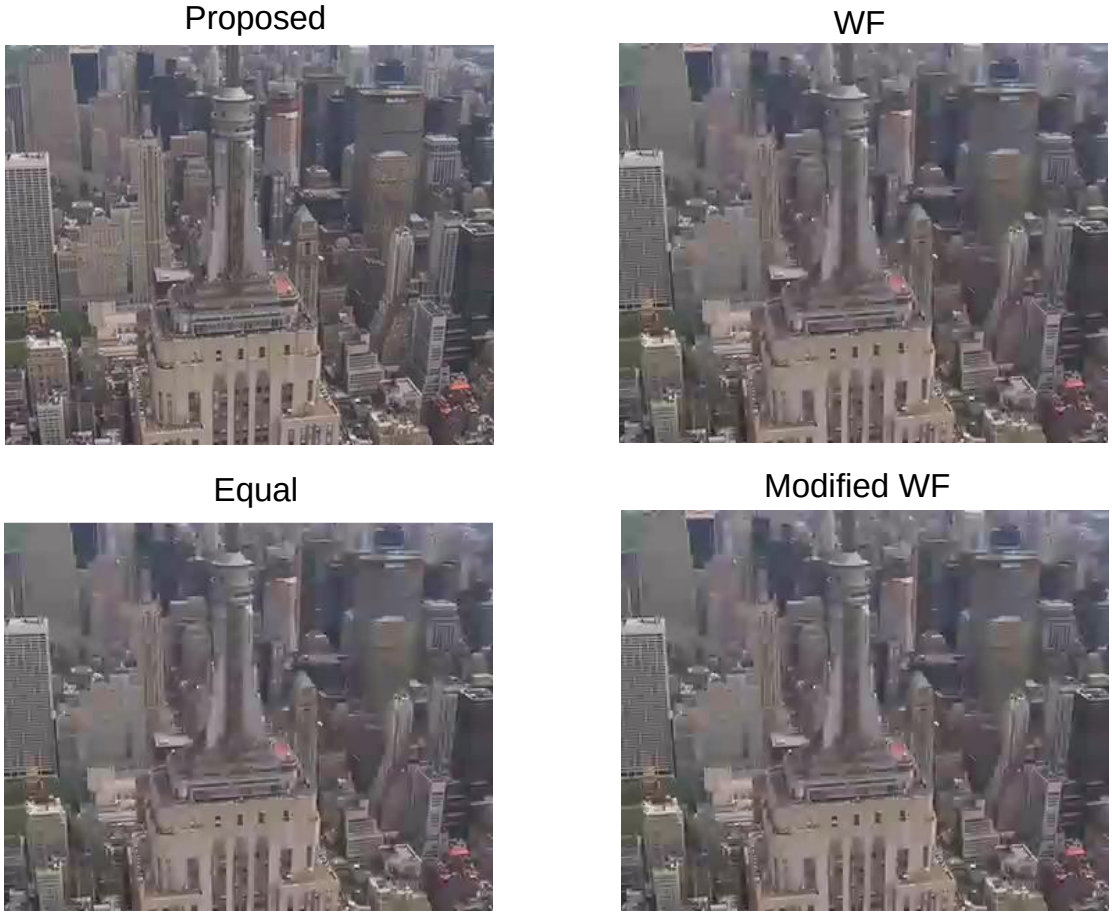

Figure 3.9: Decoded sample video frames of reconstructed video "City" at system average SNR: 24dB. Top left: proposed A1; Top right: WF; Bottom left: equal; Bottom right: M-WF.

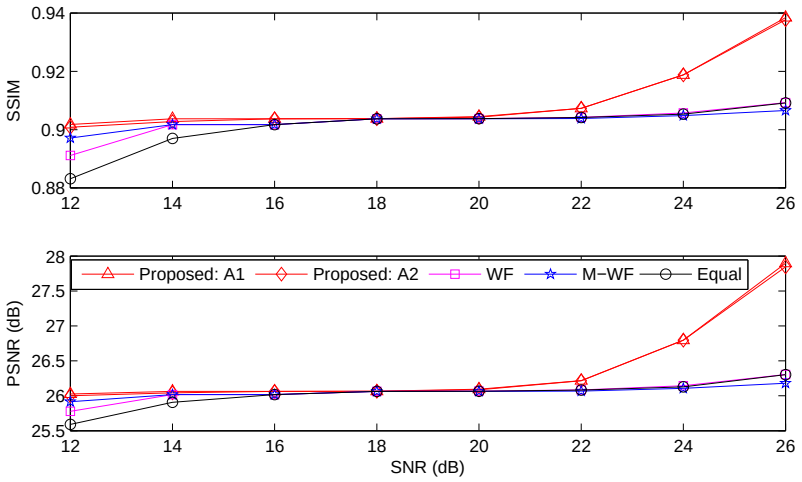

Figure 3.10: Average SSIM and PSNR of reconstructed video "City".

obvious comparing with the proposed bisection algorithm (A1). However, it is still better than other schemes.

Similar simulations can be repeated for transmitting video clip "Waterfall" with different MCSs in Table 3.1. The SVC layer indices are plotted in Fig. 3.17 where the system average SNR is set as 16dB. Still, more video frames with higher quality layers can be correctly received by using the proposed scheme. The CDF of power allocation results obtained from different schemes are shown in Fig. 3.18. Per-frame PSNR and SSIM of reconstructed video frames are illustrated in Fig. 3.19 and 3.20 respectively. The visual comparison of the reconstructed video frames is shown in Fig. 3.21. The average PSNR and SSIM with respect to system average SNR are plotted in Fig. 3.22.

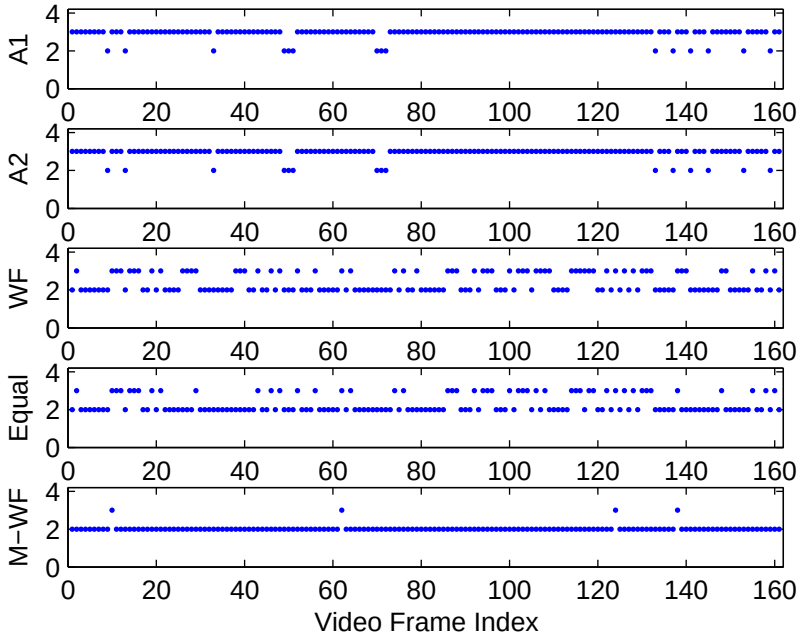

Figure 3.11: SVC quality layer indices of received frames of video "Foreman" at system average SNR: 18dB.

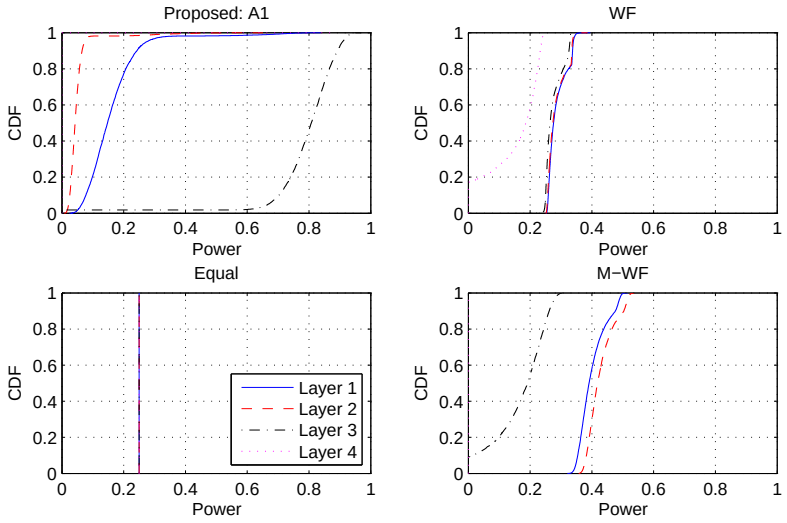

Figure 3.12: CDF of power allocation results when transmitting video "Foreman" at system average SNR: 18dB.

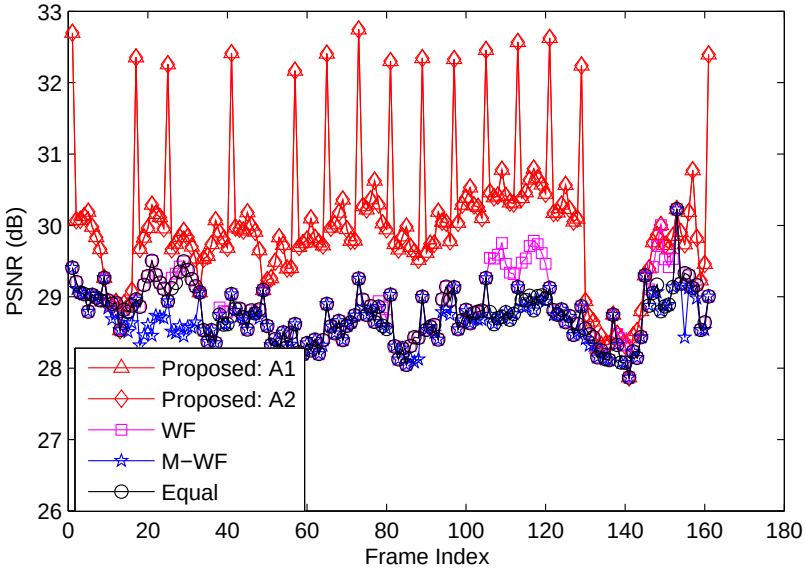

Figure 3.13: Per-frame PSNR of reconstructed video "Foreman" at system average SNR: 18dB.

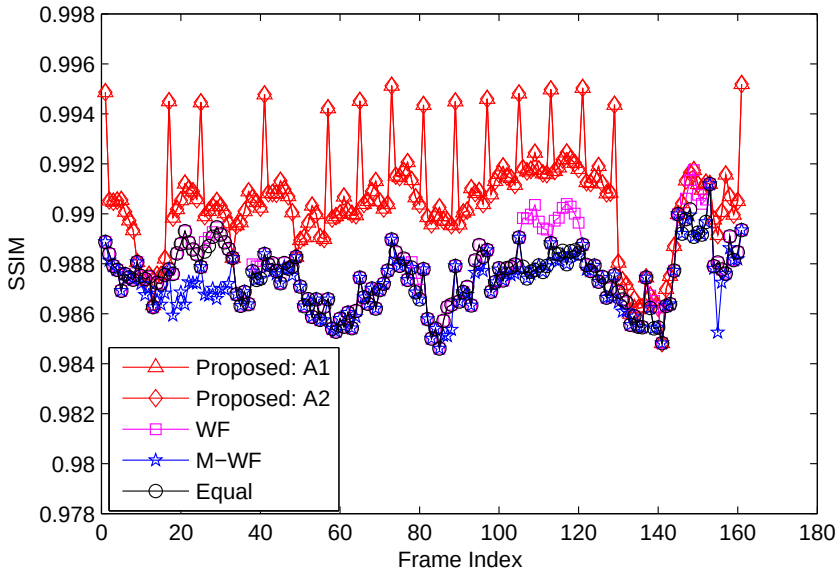

Figure 3.14: Per-frame SSIM of reconstructed video "Foreman" at system average SNR: 18dB.

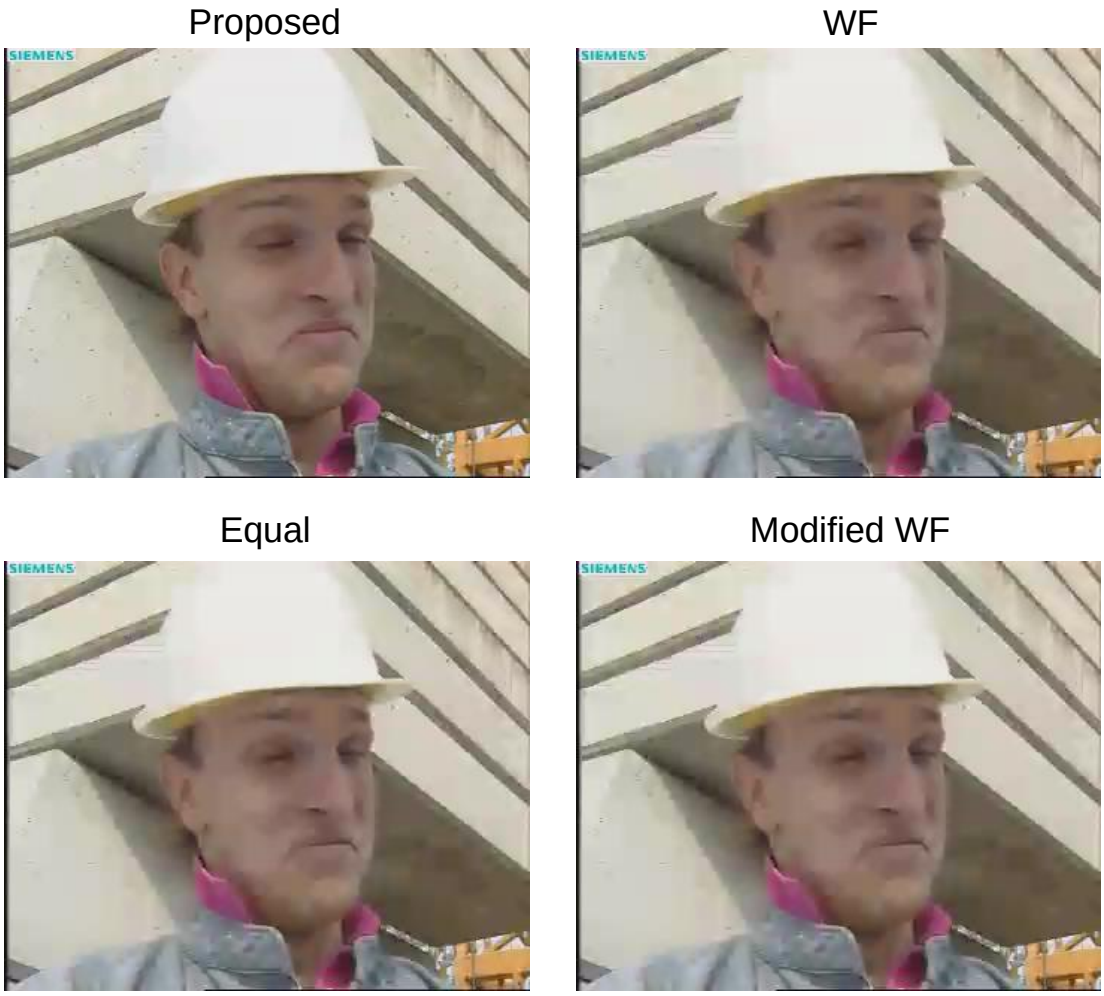

Figure 3.15: Decoded sample video frames of reconstructed video "Foreman" at system average SNR: 18dB. Top left: proposed A1; Top right: WF; Bottom left: equal; Bottom right: M-WF.

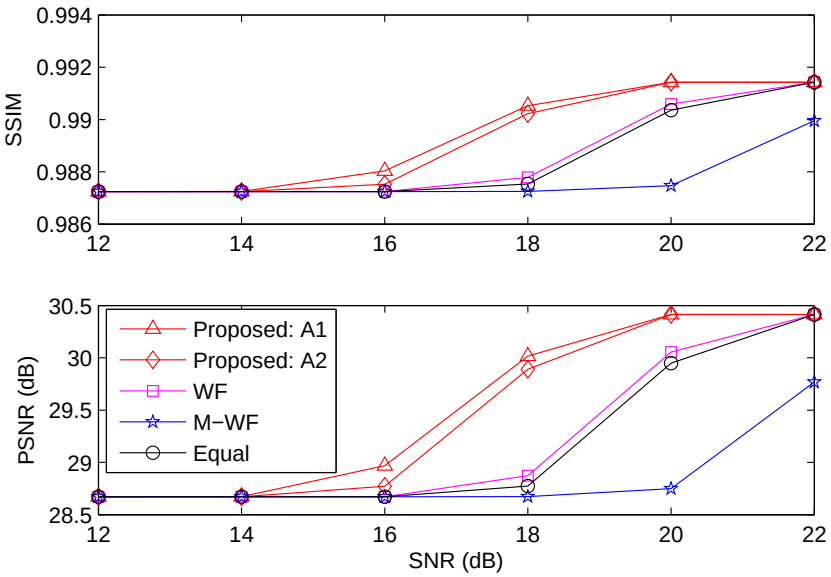

Figure 3.16: Average SSIM and PSNR of reconstructed video "Foreman".

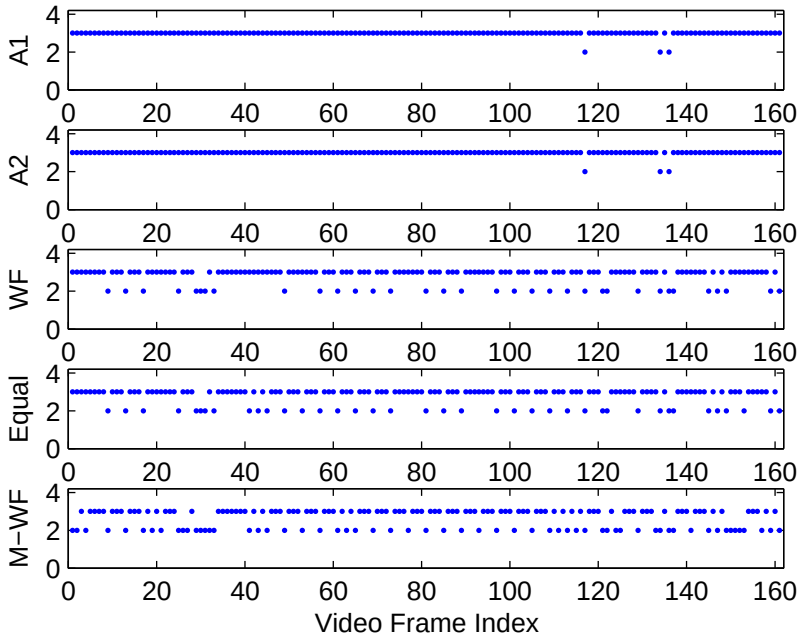

Figure 3.17: SVC quality layer indices of received frames of video "Waterfall" at system average SNR: 16dB.

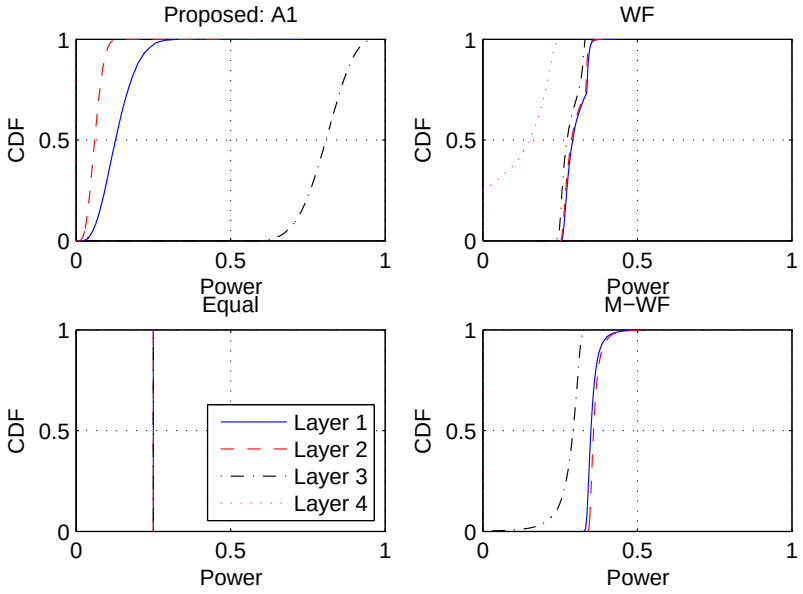

Figure 3.18: CDF of power allocation results when transmitting video "Waterfall" at system average SNR: 16dB.

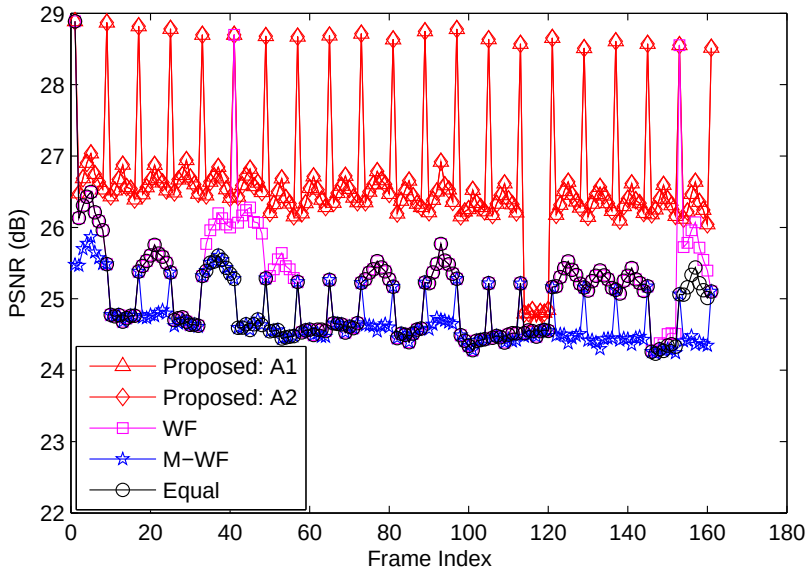

Figure 3.19: Per-frame PSNR of reconstructed video "Waterfall" at system average SNR: 16dB.

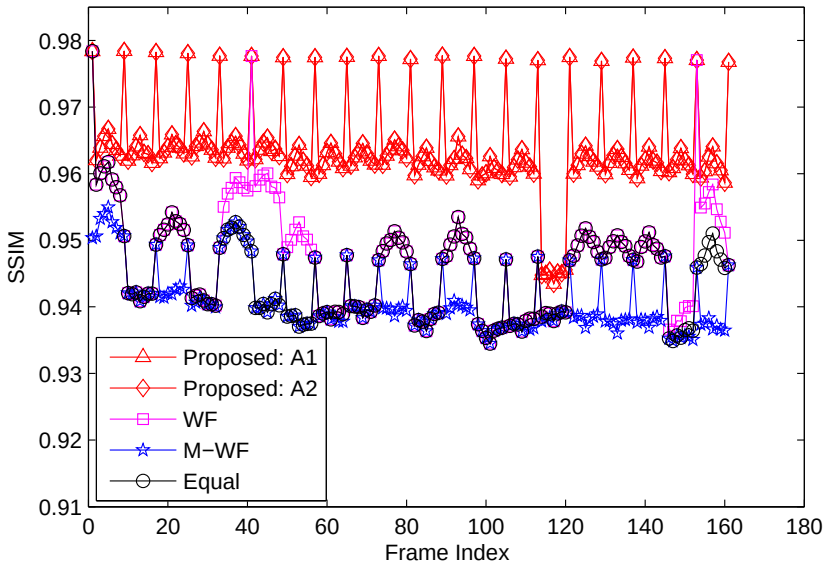

Figure 3.20: Per-frame SSIM of reconstructed video "Waterfall" at system average SNR: 16dB.

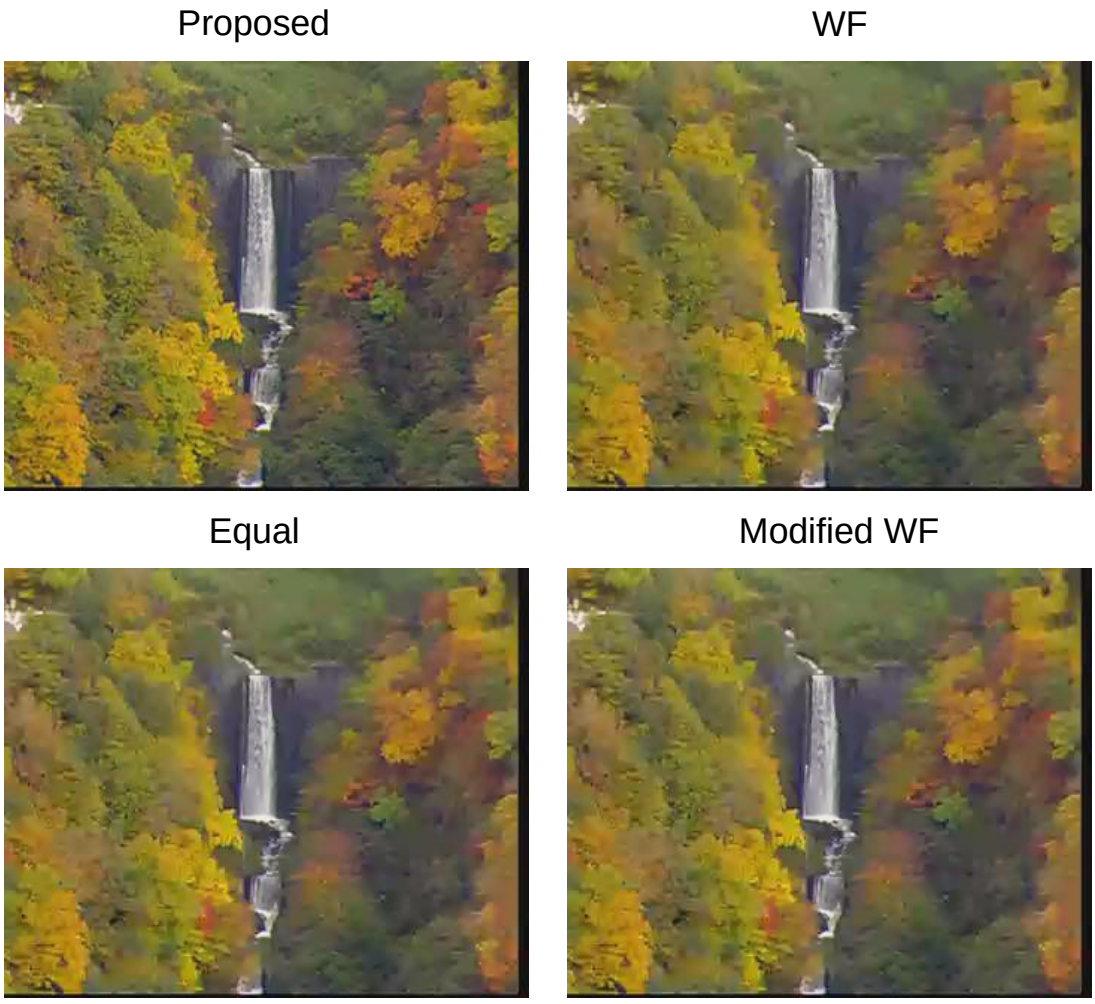

Figure 3.21: Decoded sample video frames of reconstructed video "Waterfall" at system average SNR: 16dB. Top left: proposed A1; Top right: WF; Bottom left: equal; Bottom right: M-WF.

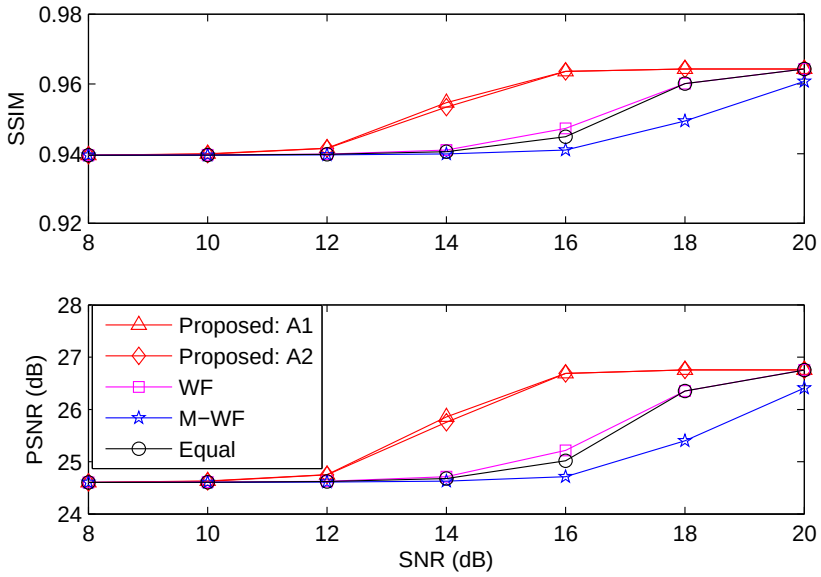

Figure 3.22: Average SSIM and PSNR of reconstructed video “Waterfall”.

Chapter 4

QUALITY-DRIVEN JOINT RATE AND POWER ADAPTATION FOR SCALABLE VIDEO TRANSMISSIONS OVER MIMO SYSTEMS

This chapter describes the proposed joint rate and power adaptation scheme to maximize the decoding quality for SVC-based video transmissions over MIMO systems. The rate adaptation in the proposed scheme includes selecting the best MCSs, set of spatial channels, number of SVC layers (source coding rates) and their corresponding APP-FEC coding rates. The power adaptation involves the proper allocation of the power to each antenna in the MIMO system. In the work described in Section 3 and most of the existing work in this field, the bit stream of each particular SVC layer is allocated to one spatial channel and the UEP is achieved by transmitting the more important SVC layers through the spatial channels with higher channel gains. However, transmitting a particular SVC layer through only one spatial channel is a waste of the limited wireless resources since the spatial diversity gain cannot be exploited. In this chapter, the bit stream of each particular SVC layer is distributed to multiple spatial channels so that additional diversity gain can be exploited by applying APP-FEC. The UEP can also be achieved by allocating different APP-FEC coding rate on each video layer. Moreover, transmit power allocation is also effectively and jointly determined to improve the system performance. To reduce the complexity of the joint rate and power adaptation problem, the original problem is decomposed into several sub-problems. Each sub-problem can be efficiently solved by series of algorithms including packet mapping, APP-FEC rate adaptation and an closed-form power allocation equation. The effectiveness and favorable performance of the proposed scheme are shown by simulations with H.264 SVC traces of HD video clips over MIMO systems.

4.1 System Overview

4.1.1 Utility and Video Decoding Quality Measure

The proposed scheme maximizes the overall video decoding quality at the receiver, which can be measured in utility values [30]. Much research has been conducted in objective video quality assessment (VQA) methods to quantify the video decoding quality. Among these VQA schemes, the Multi-Scale-Structural SIMilarity index (MS-SSIM) developed in [76], originally designed as an image quality assessment method, has been reported as a simple but effective VQA scheme, which has high correlation with subjectively perceived video quality [20, 51, 73]. In VQA applications, the MS-SSIM is applied frame-by-frame to the luminance component of the video and the overall MS-SSIM index can be calculated by averaging the frame-level quality score [73]. Therefore, the average MS-SSIM as the utility value of each encoded SVC video layer is adopted.

4.1.2 APP-FEC

Retransmission-based protocols/technologies, such as transmission control protocol (TCP) and automatic repeat request (ARQ), can provide reliable transmissions with the price of higher transmission delay. APP-FEC can provide reliable end-to-end streaming applications with packet-level protection [48], which is an alternative technology for delay-sensitive wireless video streaming applications without retransmissions [12, 13]. A source block consisting of K packets is encoded into an FEC block with N ($N \geq K$) packets so that $N-K$ redundancy packets are constructed and the encoding rate is K/N. For an ideal APP-FEC scheme, the decoder can reconstruct the original K source packets from any K out of N received packets [48] with correction capability $t = N - K$. If the packet losses are independent and identical distributed (*i.i.d.*), the APP-FEC block correction rate (BCR) of an (N, K) APP-FEC code can be expressed as [18, 30]:

$$P^{(\mathrm{BCR})}(t) = \sum_{i=0}^{t} \binom{N}{i} \left(P^{(\mathrm{PLR})}\right)^{i} \left(1 - P^{(\mathrm{PLR})}\right)^{N-i}, \tag{4.1}$$

where $P^{(\mathrm{PLR})}$ is the packet loss rate (PLR). One of the well-known APP-FEC schemes is the Reed-Solomon (RS) code, which operates on non-binary symbols and has ideal correction

capability. For the most commonly used RS code with 8 bits per symbol, at most 255 packets can be encoded in a source block, i.e., $K \leq N \leq 255$, which limits the parameters selection in practice. Moreover, the RS code has high decoding complexity due to non-binary operations [48], which is not attractive for HD video streaming applications. The Raptor code [65] is a more attractive solution for HD video streaming services due to the flexible parameters selection and linear decoding cost. Unlike RS code, the correction capability of a Raptor code is $t = N - (1 + \epsilon)K$, where ϵ is reception overhead efficiency, which makes the correction capability sub-optimal. However, the reception overhead efficiency of standardized Raptor code is close to ideal [48]. In this chapter, the APP-FEC scheme with ideal correction capability is considered.

4.1.3 MIMO System Model

The input-output relation of an $M_r \times M_t$ MIMO system can be expressed by the following linear equation:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}, \tag{4.2}$$

where $\mathbf{y}$ is an $M_r \times 1$ receive complex symbol vector. $\mathbf{H}$ is an $M_r \times M_t$ channel matrix in which all elements are *i.i.d.* zero mean circularly symmetric complex Gaussian (ZMCSCG) random variables with zero mean and unit variance, i.e., $\mathcal{CN}\{0, 1\}$. $\mathbf{x}$ is an $M_t \times 1$ transmit complex symbol vector with $E\left[\mathbf{x}\mathbf{x}^H\right] = \text{diag}\left(\mathbf{p}\right) = \text{diag}\left(p_1, p_2, ..., p_{M_t}\right)$. Here, $\mathbf{p}$ is the $M_t \times 1$ real non-negative transmission power vector subject to the normalized power constrain: $\mathbf{p}^T\mathbf{1}_{M_t} = 1$. $\mathbf{n}$ is an $M_r \times 1$ *i.i.d.* complex additive white Gaussian noise (AWGN) noise vector with covariance matrix $N_0\mathbf{I}_{M_r}$. Therefore, the system signal-to-noise ratio (SNR) is $\rho = 1/N_0$.

With a singular value decomposition (SVD), the known MIMO channel matrix $\mathbf{H}$ can be decomposed as:

$$\mathbf{H} = \mathbf{U}\mathbf{D}\mathbf{V}^H, \tag{4.3}$$

where $\mathbf{U}$ and $\mathbf{V}$ are unitary matrices. The diagonal matrix $\mathbf{D}$ is:

$$\mathbf{D} = \text{diag}\left[\sqrt{\lambda_1}, \sqrt{\lambda_2}, ..., \sqrt{\lambda_R}, 0, ..., 0\right], \tag{4.4}$$

where $R \leq \min(M_r, M_t)$ is the rank of channel matrix $\mathbf{H}$. $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_R$ are the eigenvalues of $\mathbf{H}^H\mathbf{H}$. If accurate and full channel knowledge is available at both transmitter and receiver sides, a precoder $\mathbf{V}$ and a decoder $\mathbf{U}^H$ can be appended on the transmitter and receiver side respectively so that the MIMO input-output relation can be reduced as:

$$\tilde{\mathbf{y}} = \mathbf{U}^H\mathbf{H}\mathbf{V} + \mathbf{U}^H\mathbf{n} = \mathbf{D}\mathbf{x} + \tilde{\mathbf{n}}. \tag{4.5}$$

Thus, the MIMO system can be decomposed into R independent single-input single-output (SISO) channels. The SNR of the r^{th} SISO channel is: $\text{SNR}_r = \rho\lambda_r p_r$.

4.1.4 Modulation and Coding Schemes (MCSs)

In real world wireless services, different MCSs are used for varying wireless channel conditions. In this paper, six widely used MCS schemes are adopted and their bit error rate (BER) expressions can be approximated by [49]:

$$P^{(\text{BER})}(\text{SNR}; m) \approx a_m \exp(-b_m \times \text{SNR}), \tag{4.6}$$

where a_m and b_m are coefficients corresponding to the MCS m. The coefficients of all the six MCSs are listed in Table 4.1.

4.1.5 Proposed System Structure

The system structure of the proposed cross-layer joint rate and power allocation scheme is shown in Fig. 4.1. The source video clip is encoded into $L_{\max}$ SVC layers, including one base layer and $L_{\max} - 1$ enhancement layers. The source data rate of the l^{th} SVC layer is $R_l^{(\text{S})}$. A video layer selection module chooses a suitable number of video layers, i.e., $L \leq L_{\max}$ to be transmitted. Then, the bit streams of selected video layers are packetized with fixed packet length S, which is chosen to be $S = 4800$ bits in this paper. The source packets are then fed into the APP-FEC encoder every T_s second to form an FEC block. T_s is assumed to be 10 ms in this work. To fit the source data rate of video layer l, the number of source packets per FEC block is approximately $K_l = \lceil R_l^{(\text{S})} T_s / S \rceil$. A channel selection module determines a set of spatial channels $\mathcal{S} \subseteq \mathcal{R}$ to be used for transmission, where $\mathcal{R} = \{1, 2, ..., R\}$ denotes the set of indices of available spatial channels and $|\mathcal{S}| \leq R$.

Table 4.1: Coefficients for Different MCSs [49]

Type (m)	**Modulation**	**Code Rate**	a_m	b_m	**Spectral Efficiency** (c_m)
1	BPSK	1/2	1.1369	7.5556	0.5 b/Hz
2	QPSK	1/2	0.3351	3.2543	1 b/Hz
3	QPSK	3/4	0.2197	1.5244	1.5 b/Hz
4	16 QAM	9/16	0.2081	0.6250	2.25 b/Hz
5	16 QAM	3/4	0.1936	0.3484	3 b/Hz
6	64 QAM	3/4	0.1887	0.0871	4.5 b/Hz

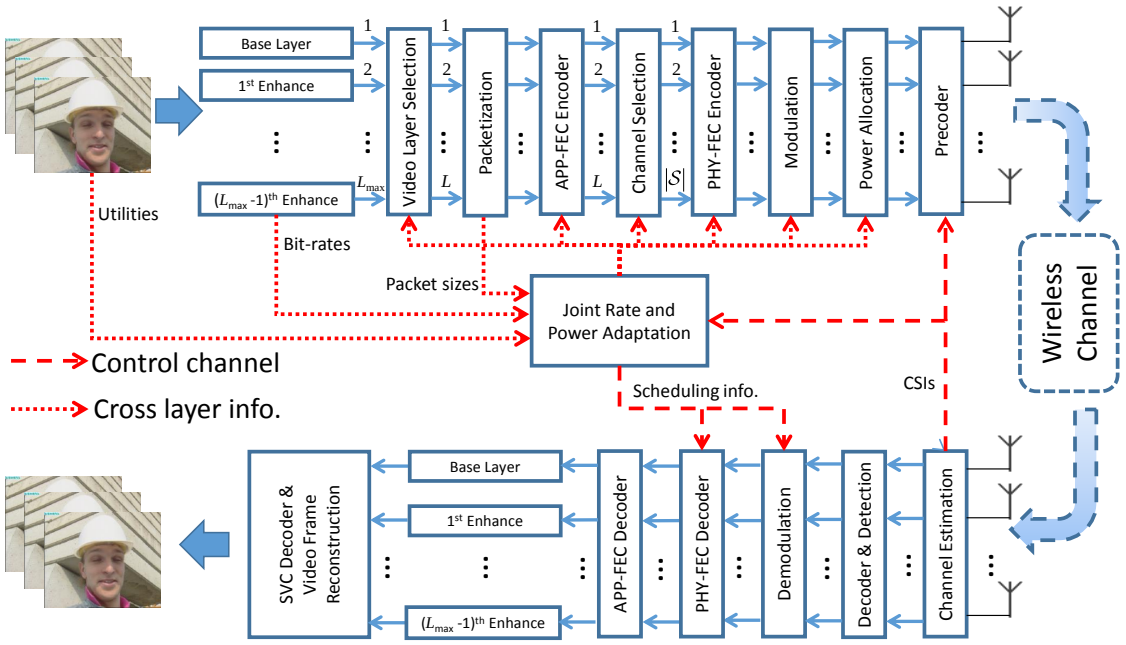

Figure 4.1: Proposed system structure for transmitting SVC-based videos to the intended receiver.

Moreover, as shown in Fig. 4.3, the channel selection module interleaves the packets of each video layer to the selected multiple spatial channels, instead of allocating the packets of more important video layers to the channels with higher channel gains as in the previous work (illustrated in Fig. 4.2). There are two advantages of the proposed channel selection method. First, the number of video layers is not restricted by the number of spatial channels because of interleaving. Second, by interleaving the packets of each video layer to several spatial channels, the diversity gain of both time and space domains can both be exploited with APP-FEC. In the proposed system, packet allocation is based on the throughput of each spatial channel, which is determined by the corresponding MCS used for transmission. For instance, if the throughput of a spatial channel r is $R_r^{(\mathrm{T})}$, $r \in \mathcal{S}$, and in total $N^{(l)}$ encoded packets of a video layer l needs to be transmitted, the number of packets of the video layer l transmitted through the spatial channel r is approximately

$$N_{r,l} \approx \frac{N^{(l)} R_r^{(\mathrm{T})}}{\sum_{k \in \mathcal{S}} R_k^{(\mathrm{T})}}. \tag{4.7}$$

The exact number of packets can be calculated by the algorithm described in Section 4.3. The modulated symbols are transmitted through the wireless channel after power allocation and precoding. All the transmission parameters, including number of SVC layers (source coding rates), APP-FEC rate of each video layer, MCS of each spatial channel and its corresponding transmit power, are jointly optimized by the proposed joint rate and power adaptation scheme based on cross-layer information at the transmitter side and channel state information (CSI) from the receiver side. At the receiver, the channel estimation module estimates the channel and feeds back the CSI. In this work, perfect and full channel knowledge at both transmitter and receiver sides is assumed. After MIMO decoding, detection, demodulation, PHY-FEC decoding and APP-FEC decoding, the received bit streams are saved in receiving buffers and fed into the SVC decoder. The bit stream of the video layer l is dropped if a transmission error is detected or its dependent layers (i.e., from the base layer to the $\left(l - 1^{\mathrm{th}}\right)$ layer) are not correctly decoded.

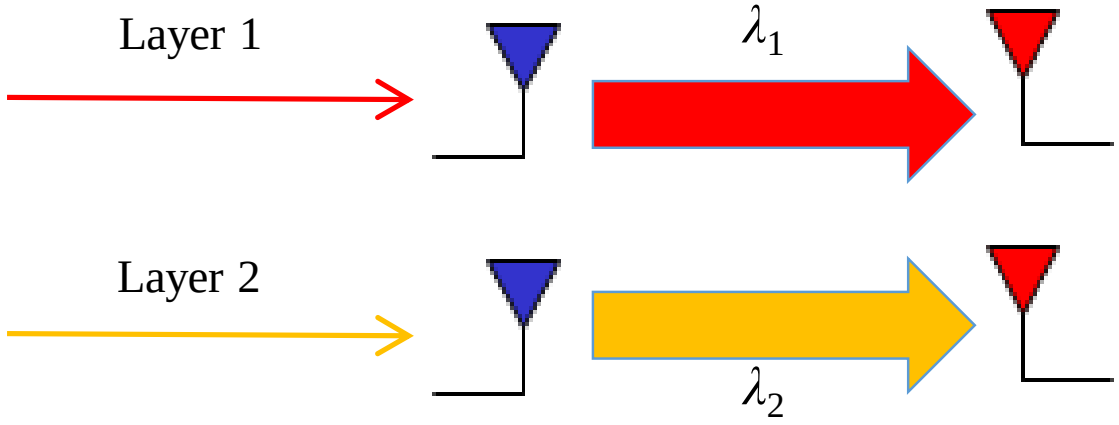

Figure 4.2: Concept of packet stream allocation in the previous work. The more important packet stream is transmitted through the spatial channels with higher gains

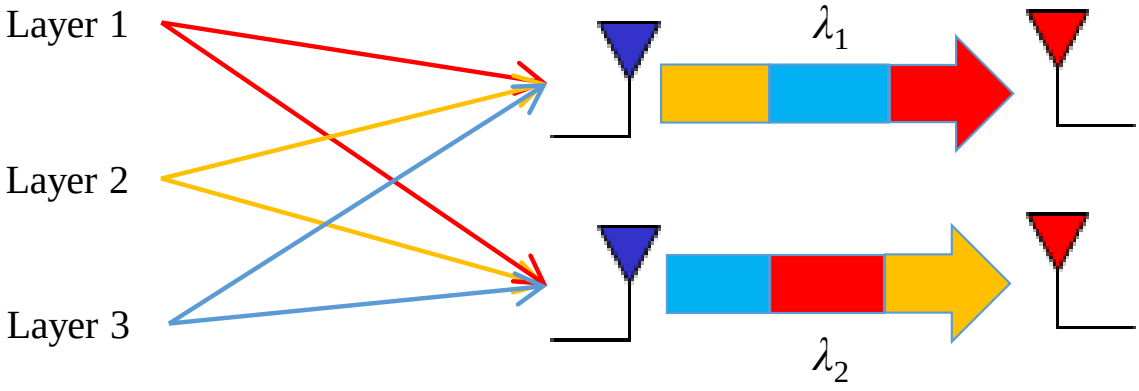

Figure 4.3: Concept of packet stream interleaving performed in the proposed channel selection module. The packet stream of each video layer is interleaved to multiple spatial channels.

4.2 Problem Formulation

The video decoding quality at the receiver side can be optimized by maximizing the average utility of the system. The optimization problem is defined as:

$$\begin{aligned} \mathbb{Q}: \quad & \max_{L,\mathbf{m},\mathbf{t},\mathbf{p}} \quad \sum_{l=1}^{L} u_l P_l^{(\mathrm{V})}(\mathbf{m},\mathbf{t},\mathbf{p}) \\ & \text{subject to} \quad \mathbf{p} \succeq 0, \\ & \qquad \sum_{r=1}^{R} p_r = 1, \\ & \qquad \mathbf{t} \succeq 0, \\ & \qquad \sum_{l=1}^{L} t_l + K_l = \sum_{r=1}^{R} \lfloor T_s R_r^{(\mathrm{T})}(m_r)/S \rfloor, \end{aligned} \tag{4.8}$$

where L is the number of SVC video layers selected for transmissions. $\mathbf{t}$ is an $L \times 1$ vector and its l^{th} element $t_l \triangleq [\mathbf{t}]_l$ denotes the APP-FEC correction capability of the SVC layer l. $\mathbf{m} \triangleq [m_1, m_2, ..., m_R]^T$, $m_r = 0, 1, ..., 6$, denote the MCSs of the r^{th} spatial channel with corresponding spectral efficiency c_{m_r} in Table 4.1. Here $m_r = 0$ indicates that the r^{th} channel is not selected for transmission (i.e., $r \notin \mathcal{S}$). $\mathbf{p}$ is an $R \times 1$ vector and its r^{th} element $p_r \triangleq [\mathbf{p}]_r$ is the transmit power allocated on the r^{th} channel. u_l is the utility value associated with the l^{th} SVC layer. $R_r^{(\mathrm{T})}(m_r)$ is the throughput of spatial channel r, which is a function of the corresponding chosen MCS. $P_l^{(\mathrm{V})}(\mathbf{m},\mathbf{t},\mathbf{p})$ is the probability of the successfully decoding SVC layer l, and can be represented as:

$$P_l^{(\mathrm{V})}(\mathbf{m},\mathbf{t},\mathbf{p}) = \prod_{j=1}^{l} \left(P_j^{(\mathrm{BCR})}(\mathbf{m}, t_j, \mathbf{p}) \right)^b, \tag{4.9}$$

where b denotes the number of APP-FEC blocks in each group of pictures (GoP) time period. $P_j^{(\mathrm{BCR})}(\mathbf{m}, t_j, \mathbf{p})$ is the APP-FEC block correction rate of the j^{th} SVC layer. Define $X_{r,j}$ as a random variable of having erroneous packets in one APP-FEC block of the SVC layer j transmitted through spatial channel r, it follows the Binomial distribution and its probability mass function (PMF) is:

$$f_{X_{r,j}}(x) = \binom{N_{r,j}}{x} \left(P_r^{(\mathrm{PLR})} \right)^x \left(1 - P_r^{(\mathrm{PLR})} \right)^{N_{r,j}-x}, \tag{4.10}$$

where the PLR of the r^{th} spatial channel is given by:

$$\begin{aligned} P_r^{(\text{PLR})} &= 1 - \left(1 - P_r^{(\text{BER})}\left(p_r, m_r\right)\right)^S \\ &= 1 - \left(1 - a_{m_r} \exp\left(-b_{m_r} \rho \lambda_r p_r\right)\right)^S. \end{aligned} \tag{4.11}$$

Thus, the APP-FEC block correction rate in (4.9) can be represented as:

$$P_j^{(\text{BCR})}\left(\mathbf{m}, t_j, \mathbf{p}\right) = \sum_{k=0}^{t_j} f_{Y_j}\left(k\right), \tag{4.12}$$

where $f_{Y_j}\left(y\right)$ is the PMF of random variable Y_j and

$$Y_j = \sum_{r \in \mathcal{S}} X_{r,j}. \tag{4.13}$$

Therefore, $f_{Y_j}\left(y\right)$ can be derived by calculating the convolution sums of $f_{X_{r,j}}\left(x\right)$. In (4.8), the first constraint means that the transmit power allocated on each spatial channel is non-negative. The second constraint means that the sum of power is limited by a unit power 1. The third constraint means that the error correction capability of each APP-FEC block is non-negative. The last constraint sets the total number of transmitted packets in each sampling period of an APP-FEC encoding block (i.e., T_s seconds) equal to the total number of packets that the network can transmit in this time period.

4.3 Proposed Algorithms

4.3.1 Sub-Problems

To reduce the complexity of the original optimization problem in (4.8), Let the maximum SVC layers to be L, with the MCS vector being $\mathbf{m}$, and decompose the original problem into several sub-problems:

$$\begin{aligned} \mathbb{Q}_{L,C}: \quad & \max_{\mathbf{t},\mathbf{p}} \quad P_L^{(\text{V})}\left(\mathbf{t}, \mathbf{p}; \mathbf{m}^{(C)}\right) \\ & \text{subject to} \quad \mathbf{p} \succeq 0, \\ & \qquad \sum_{r=1}^{R} p_r = 1, \\ & \qquad \mathbf{t} \succeq 0, \\ & \qquad \sum_{l=1}^{L} t_l + K_l = \sum_{r=1}^{R} \left\lfloor \frac{T_s R_r^{(\text{T})}\left(m_r^{(C)}\right)}{S} \right\rfloor, \end{aligned} \tag{4.14}$$

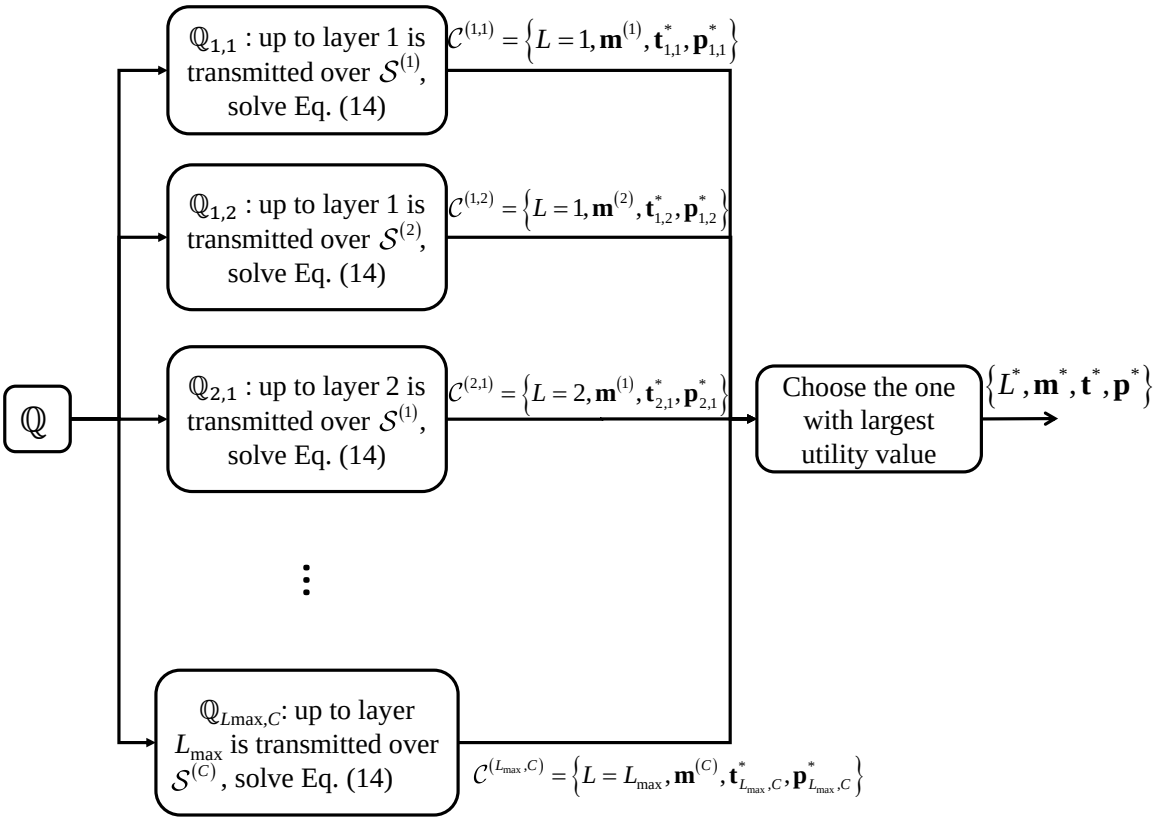

Figure 4.4: Problem solving strategies: decomposing the original problem into several sub-problems.

where $\mathbf{m}^{(C)}$ is the C^{th} possible combination of the MCS vector $\mathbf{m}$. Each sub-problem provides one candidate solution set $\left\{L, \mathbf{m}^{(C)}, \mathbf{t}^*_{L,C}, \mathbf{p}^*_{L,C}\right\}$. Note that for each sub-problem, the parameters L and $\mathbf{m}^{(C)}$ are pre-determined. The solution of the original problem can be obtained by searching the candidate solution set with highest utility, i.e.,

$$\{L^*, \mathbf{m}^*, \mathbf{t}^*, \mathbf{p}^*\} = \arg\max_{\mathcal{C}} \sum_{l=1}^{L} u_l P_l^{(\mathrm{V})} \left(\mathbf{t}^*_{L,C}, \mathbf{p}^*_{L,C}; \mathbf{m}^{(C)}\right), \tag{4.15}$$

where $\mathcal{C}$ denotes the set containing all $L_{\max} \times C$ candidate solutions. The problem solving strategies including the relationships between the original problem, the sub-problems, the candidate solution set and the final solution are illustrated in Fig. 4.4.

Figure 4.5 shows the proposed modules for each sub-problem. A weighting vector is estimated first. Then, the estimated weightings and necessary parameters are fed into the

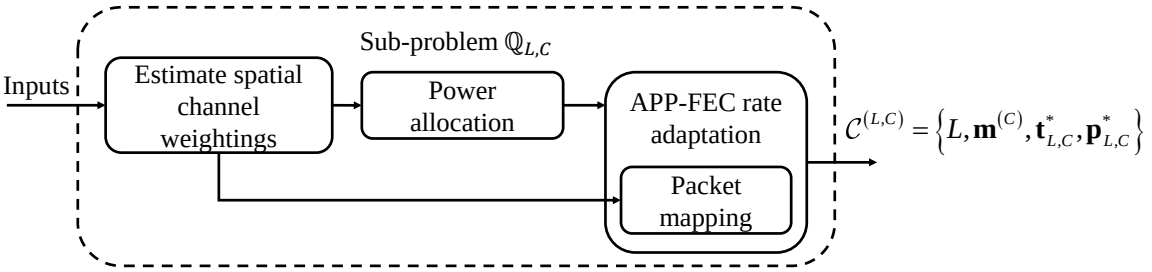

Figure 4.5: Proposed modules for each sub-problem.

power allocation module and the packet mapping module. The power allocation module calculates the optimal power and is used by the APP-FEC rate adaptation module, which also includes packet mapping (interleaving) to each spatial channel.

4.3.2 Spatial Channel Weightings and Packet Mapping

To exploit the diversity gain of different spatial channels, the APP-FEC encoded packets of each SVC layer are mapped to multiple spatial channels. As indicated in (4.7), the number of packets of the l^{th} SVC layer over the r^{th} channel (i.e., $N_{r,l}$) is proportional to the normalized throughput of the r^{th} channel. Let $\mathbf{N}$ be an $R \times L$ packet mapping matrix with elements $N_{r,l}$. Let $\mathbf{w}$ be an $R \times 1$ vector with each element

$$w_r = \frac{R_r^{(\mathrm{T})}}{\sum_{k \subseteq \mathcal{S}} R_k^{(\mathrm{T})}} \tag{4.16}$$

representing the rate distribution target (weighting) of the r^{th} spatial channel. The throughput of the r^{th} channel is

$$R_r^{(\mathrm{T})} = B c_{m_r}. \tag{4.17}$$

Table 4.2 shows the packet mapping algorithm to calculate $\mathbf{N}$.

The proposed packet mapping algorithm iteratively allocates one packet to the spatial channel whose weighting $w_r^{(\text{temp})}$, calculated in line 4 of Table 4.2, is the smallest comparing to its corresponding target weighting w_r. And the packet number distribution for each spatial channel r approaches to the target weighting w_r for each iteration.

Table 4.2: Algorithm for packet mapping

Inputs:	Weighting vector: $\mathbf{w}$
	Packet number: $N^{(l)}$ for each $l = 1, 2, ..., L$
Output:	Packet mapping matrix: $\mathbf{N}$
1.	Initialization: $N_{r,l} := 0$ for all r and l; $N_{1,1} := 1$;
2.	**for** $l = 1 \rightarrow L$
3.	**while** $N^{(l)} > 0$
4.	$w_r^{(\text{temp})} := \sum_l N_{r,l} / \sum_{r,l} N_{r,l}$ for each $r \in \mathcal{S}$;
5.	$r^* := \arg\max \left(w_r - w_r^{(\text{temp})} \right)$ for each $r \in \mathcal{S}$;
6.	$N_{r^*,l} := N_{r^*,l} + 1$;
7.	$N^{(l)} := N^{(l)} - 1$;
8.	**end while**
9.	**end for**
10.	**return N**;

4.3.3 Power Allocation

As indicated in (4.14), the goal of power allocation is to maximize the successful decoding probability $P_L^{(\mathrm{V})}\left(\mathbf{t},\mathbf{p};\mathbf{m}^{(C)}\right)$, which is equivalent to maximizing the APP-FEC block correction rate $P_j^{(\mathrm{BCR})}\left(t_j,\mathbf{p};\mathbf{m}\right)$ in (4.9). When $N_{r,j}$ is large and $P_r^{(\mathrm{PLR})}$ is small, the PMF of $X_{r,j}$ can be approximated by the Poisson distribution with parameter $\gamma_{X_{r,j}} = N_{r,j} \times P_r^{(\mathrm{PLR})}$. Therefore, Y_j in (4.13) also follows Poisson distribution with parameter [71]

$$\gamma_{Y_j}\left(\mathbf{p}\right) = \sum_{r\in\mathcal{S}} N_{r,j} P_r^{(\mathrm{PLR})}\left(p_r\right) \approx N^{(j)} \sum_{r\in\mathcal{S}} w_r P_r^{(\mathrm{PLR})}\left(p_r\right). \tag{4.18}$$

If the MCS vector $\mathbf{m}$ and the APP-FEC correction capability $\mathbf{t}$ are given, the APP-FEC block correction rate $P_j^{(\mathrm{BCR})}\left(\mathbf{p};\mathbf{m},t_j\right)$ is maximized when $\gamma_{Y_j}\left(\mathbf{p}\right)$ is minimized. Therefore the power allocation problem can be written as

$$\begin{aligned} \mathbb{Q}_{L,C}^{(\mathrm{PA})}: \quad & \mathbf{p}_{L,C}^{*} = \arg\min_{\mathbf{p}} J\left(\mathbf{p};\mathbf{m}^{(C)}\right) \\ \text{subject to} \quad & \mathbf{p} \succeq 0, \\ & \sum_{r=1}^{R} p_r = 1, \end{aligned} \tag{4.19}$$

where

$$J\left(\mathbf{p};\mathbf{m}^{(\mathrm{C})}\right) = \sum_{r\in\mathcal{S}^{(C)}} w_r\left(m_r^{(C)}\right) P_r^{(\mathrm{PLR})}\left(p_r\right) \tag{4.20}$$

is the weighted sum of packet loss rate of each spatial channel r.

When $P_r^{(\mathrm{BER})}$ is small, $P_r^{(\mathrm{PLR})}$ in (4.11) can be approximated by

$$P_r^{(\mathrm{PLR})}\left(p_r;m_r\right) \approx S \cdot a_{m_r} \exp\left(-b_{m_r}\rho\lambda_r p_r\right), \tag{4.21}$$

which is convex function with respect to p_r. Therefore, the objective function in (4.20) can also be approximated as a convex function since the non-negative sum of convex functions is still convex [5]. The power allocation problem in (4.19) can thus be efficiently solved by (4.22) when the approximation is valid (see Appendix A for derivations).

$$\left[\mathbf{p}_{L,C}^{*}\right]_r = \begin{cases} \dfrac{\mu + \log\left(S w_r a_{m_r} b_{m_r} \rho\lambda_r\right)}{b_{m_r}\rho\lambda_r}, & r \in \mathcal{S} \\ 0, & r \notin \mathcal{S} \end{cases} \tag{4.22}$$

where

$$\mu = \frac{1 - \sum_{r\in\mathcal{S}} \frac{\log(Sw_r a_{m_r} b_{m_r}\rho\lambda_r)}{b_{m_r}\rho\lambda_r}}{\sum_{r\in\mathcal{S}} (1/b_{m_r}\rho\lambda_r)}. \tag{4.23}$$

However, when the approximation is not valid, i.e., $P_r^{(\mathrm{BER})}$ is not small enough, more power than necessary is allocated on the channels with low channel gains, resulting in wasted power. Let $J_{\mathrm{appx}}(\mathbf{p})$ denotes the objective function in (4.19) with the approximation in (4.21), the power allocation solution is trustable if $\left|J\left(\mathbf{p}_{L,C}^*;\mathbf{m}^{(C)}\right) - J_{\mathrm{appx}}\left(\mathbf{p}_{L,C}^*;\mathbf{m}^{(C)}\right)\right|$ is less than a threshold Th, which is empirically set as 0.001. Otherwise, the channel with smallest gain is removed from the selected channel set $\mathcal{S}$ and its corresponding power is set as 0. Then, $p_{L,C}^*$ is recalculated by (4.22) until the solution is trustable or only one channel is left in $\mathcal{S}$, i.e., $|\mathcal{S}| = 1$. The algorithm for power allocation is described in Table 4.3.

4.3.4 APP-FEC Rate Adaptation

Since the transmitted packets of each SVC layer are spread over multiple spatial channels, the channel conditions experienced by each SVC layer are similar. Therefore, the UEP cannot be achieved by channel selection as in [10]. Instead, under the overall allowed packet number constraint, determined by the system throughput, the UEP can be achieved by assigning different APP-FEC rates on each SVC layer. This results in the following APP-FEC rate adaptation problem:

$$\begin{aligned} \mathbb{Q}_{L,C}^{(\mathrm{FEC})}: \quad & \mathbf{t}_{L,C}^* = \arg\max_{\mathbf{t}} P_L^{(\mathrm{V})}\left(\mathbf{t};\mathbf{m}^{(C)},\mathbf{p}_{L,C}^*\right) \\ \text{subject to} \quad & \mathbf{t} \succeq 0, \\ & \sum_{l=1}^{L} t_l + K_l = \sum_{r=1}^{R}\left\lfloor T_s R_r^{(\mathrm{T})}\left(m_r^{(C)}\right)/S \right\rfloor. \end{aligned} \tag{4.24}$$

Note that the APP-FEC rate adaptation is conducted after the optimal power allocation vector $\mathbf{p}_{L,C}^*$ is obtained. Therefore, all the parameters L, $\mathbf{m}^{(C)}$, and $\mathbf{p}_{L,C}^*$ are determined and the only variable left is $\mathbf{t}$. In this paper, a simple steepest ascent algorithm, shown in Table 4.4, is proposed to find the optimal solution of the APP-FEC rate of each SVC layer. Note that the objective function of (4.24), $P_L^{(\mathrm{V})}(\mathbf{t})$ is non-decreasing with respect to the FEC correction capabilities $\mathbf{t}$. Therefore, the steepest ascent algorithm can reach the global optimal solution.

Table 4.3: Algorithm for power allocation

Inputs:	Channel MCS vector: $\mathbf{m}^{(C)}$
	Channel gain: $\rho\lambda_r$ for each $r = 1, 2, ..., R$
Output:	Power allocation vector: $\mathbf{p}^*_{L,C}$
1.	**if** $\lvert\mathcal{S}\rvert = 1$;
2.	$\left[\mathbf{p}^*_{L,C}\right]_1 := 1$;
3.	$\left[\mathbf{p}^*_{L,C}\right]_r := 0$ for $r = 2, 3, ..., R$;
4.	**else**
5.	Solve $\mathbf{p}^*_{L,C}$ by (4.22);
6.	**if** $\left\lvert J\left(\mathbf{p}^*_{L,C}; \mathbf{m}^{(C)}\right) - J_{\text{appx}}\left(\mathbf{p}^*_{L,C}; \mathbf{m}^{(C)}\right)\right\rvert \geq Th$
7.	$r^{(\min)} := \arg\min_r \lambda_r$;
8.	$\mathcal{S} := \mathcal{S} - \left\{r^{(\min)}\right\}$;
9.	$\left[\mathbf{p}^*_{L,C}\right]_{r^{(\min)}} := 0$;
10.	Go back to step 1;
11.	**end if**
12.	**end if**
13.	**return** $\mathbf{p}^*_{L,C}$;

Table 4.4: Algorithm for APP-FEC rate adaptation

Inputs:	Number of APP-FEC blocks: b
	Number of source packets: K_l for each $l = 1, 2, ..., L$
	Channel MCS vector: $\mathbf{m}^{(C)}$
	System bandwidth: B
	APP-FEC block sampling time: T_s
	APP-FEC packet size: S
	Channel SNR gains: $\rho\lambda_r$ for each $r = 1, 2, ..., R$
	Optimal power allocations: p_r^* for each $r = 1, 2, ..., R$
Output:	APP-FEC correction capability vector: $\mathbf{t}_{L,C}^*$
1.	Initialization: $\mathbf{t} := \mathbf{0}$;
2.	**while** $1 + \sum_{l=1}^{L} t_l + K_l \leq \sum_{r=1}^{R} \left\lfloor T_s R_r^{(\mathrm{T})} \left(m_r^{(C)} \right) / S \right\rfloor$
3.	**for** $j = 1 \rightarrow L$
4.	Do packet mapping algorithm and obtain $\mathbf{N}$;
5.	$\mathbf{t}^{(\mathrm{temp})} := \mathbf{t}$;
6.	$t_j^{\mathrm{temp}} := t_j + 1$;
7.	Calculate $P_j^{(\mathrm{BCR})}$ by (4.12);
8.	$G_j(t_j) := P_L^{(\mathrm{V})} \left(\mathbf{t}^{\mathrm{temp}}; \mathbf{m}^{(C)}, \mathbf{p}_{L,C}^* \right) - P_L^{(\mathrm{V})} \left(\mathbf{t}; \mathbf{m}^{(C)}, \mathbf{p}_{L,C}^* \right)$;
9.	**end for**
10.	$j^* := \arg\max G_j(t_j)$;
11.	$t_{j^*} := t_{j^*} + 1$;
12.	**end while**
13.	**return** $\mathbf{t}_{L,C}^* := \mathbf{t}$;

The idea of the proposed APP-FEC rate adaptation algorithm iteratively adds one APP-FEC correction capability to the SVC layer with the highest gain in terms of video decoding correction probability. The iteration stops when no additional redundancy packets can be added due to the throughput limitation.

4.3.5 Overall Rate and Power Allocation Adaptation Algorithm

The overall proposed joint rate and power allocation adaptation algorithm, which combines all the above mentioned algorithms, is described in Table 4.5.

4.4 Simulation Results

The effectiveness and favorable performance of the proposed algorithm are evaluated by intensive simulations. Two high definition (HD) video clips "Cactus" and "Kimono" are encoded by the Joint Scalable Video Model (JSVM) version 9.19 [2]. The detailed video coding and transmission parameters are listed in Table 4.6. Both videos are encoded with medium-grain scalability (MGS). Both motion compensation and estimation are constrained at the current layer [10]. The cumulative source coding bit rates of the 4 SVC layers are listed in Table 4.7. The encoded network abstraction layer units (NALUs) are packetized into packets with 600 bytes. A 4×4 MIMO system is used for transmissions. The channel matrix changes randomly every channel coherence time. And the CSIs with full channel knowledge are fed back every channel coherence time. The sampling period T_s of APP-FEC block is 10 ms. At the receiver side, control messages such as video coding parameters and MCS information are assumed to be correctly received. Perfect error detection is assumed and erroneous packets are dropped. The undecodable NALU, which is caused by either packet loss of its own SVC layer packets or unsatisfied SVC layer dependencies, is discarded before SVC decoding.

In total 6 different schemes are simulated. Their corresponding techniques and abbreviations are listed in Table 4.8. Note that the "FEC" scheme, which uses APP-FEC to exploit the diversity of different spatial channels, is similar to the previous work in [9]. The "Baseline" scheme in Table 4.8 refers to a similar system proposed in [69], where the bit stream of the l^{th} SVC layer is transmitted through the spatial channel with the l^{th} highest

Table 4.5: Overall algorithm for rate and power adaptation

Inputs:	Number of APP-FEC blocks: b
	Maximum SVC layer: $L_{\max}$
	Source packet number: K_l for each $l = 1, 2, ..., L_{\max}$
	Possible MCSs table
	Utility vector: $\mathbf{u}$
	System bandwidth: B
	APP-FEC block sampling time: T_s
	APP-FEC packet size: S
	Channel gain: $\rho\lambda_r$ for each $r = 1, 2, ..., R$
Output:	Optimal solution set: $\{L^*, \mathbf{m}^*, \mathbf{t}^*, \mathbf{p}^*\}$
1.	Initialization: $\text{utility}^{(\max)} := 0$;
2.	**for** each MCS combination C
3.	Calculate weighting vector $\mathbf{w}$ by (4.16);
4.	Do power allocation and obtain $\mathbf{p}^*_{L,C}$;
5.	**for** $L = 1 \rightarrow L_{\max}$
6.	Do APP-FEC rate adaptation and obatin $\mathbf{t}^*_{L,C}$
7.	$\text{utility} := \sum_{l=1}^{L} u_l P_l^{(\mathrm{V})} \left(\mathbf{m}^{(C)}, \mathbf{t}^*_{L,C}, \mathbf{p}^*_{L,C}\right)$;
8.	**if** $\text{utility} > \text{utility}^{(\max)}$
9.	$\text{utility}^{(\max)} := \text{utility}$;
10.	$\{L^*, \mathbf{m}^*, \mathbf{t}^*, \mathbf{p}^*\} := \left\{L, \mathbf{m}^{(C)}, \mathbf{t}^*_{L,C}, \mathbf{p}^*_{L,C}\right\}$;
11.	**end if**
12.	**end for**
13.	**end for**
14.	**return** $\{L^*, \mathbf{m}^*, \mathbf{t}^*, \mathbf{p}^*\}$;

Table 4.6: Video Coding and Transmission Parameters

	Cactus	**Kimono**
GoP Size (Frames)	8	8
Frame Pattern	IBBBBBBB	IBBBBBBB
Total SVC Layers	4	4
QP of SVC Layer 1 (Base Layer)	46	46
QP of SVC Layer 2	36	36
QP of SVC Layer 3	26	26
QP of SVC Layer 4	16	16
Total Encoded Frames	497	233
Frame Rate (fps)	50	24
System Bandwidth (MHz)	15	5
Coherence Time (ms)	50	10

Table 4.7: Cumulative Bit Rate of Different Video Layers

Video	**Layer 1**	**Layer 2**	**Layer 3**	**Layer 4**
Cactus	2.487 Mbps	5.389 Mbps	16.79 Mbps	59.92 Mbps
Kimono	1.0327 Mbps	1.889 Mbps	5.844 Mbps	21.94 Mbps

Table 4.8: Proposed and Control Group Schemes

Abbreviations	**Techniques**
AMC+FEC+PA	Proposed algorithm in Table 4.5
AMC+FEC	Proposed algorithm in Table 4.5 with equal power allocation on each spatial channel.
FEC+PA	Proposed algorithm in Table 4.5 with fixed MCS on each spatial channel.
FEC [9]	Proposed algorithm in Table 4.5 with fixed MCS and equal power allocation on each spatial channel.
Baseline+PA [10]	Similar system as in [10]
Baseline [69]	Similar system as in [69]

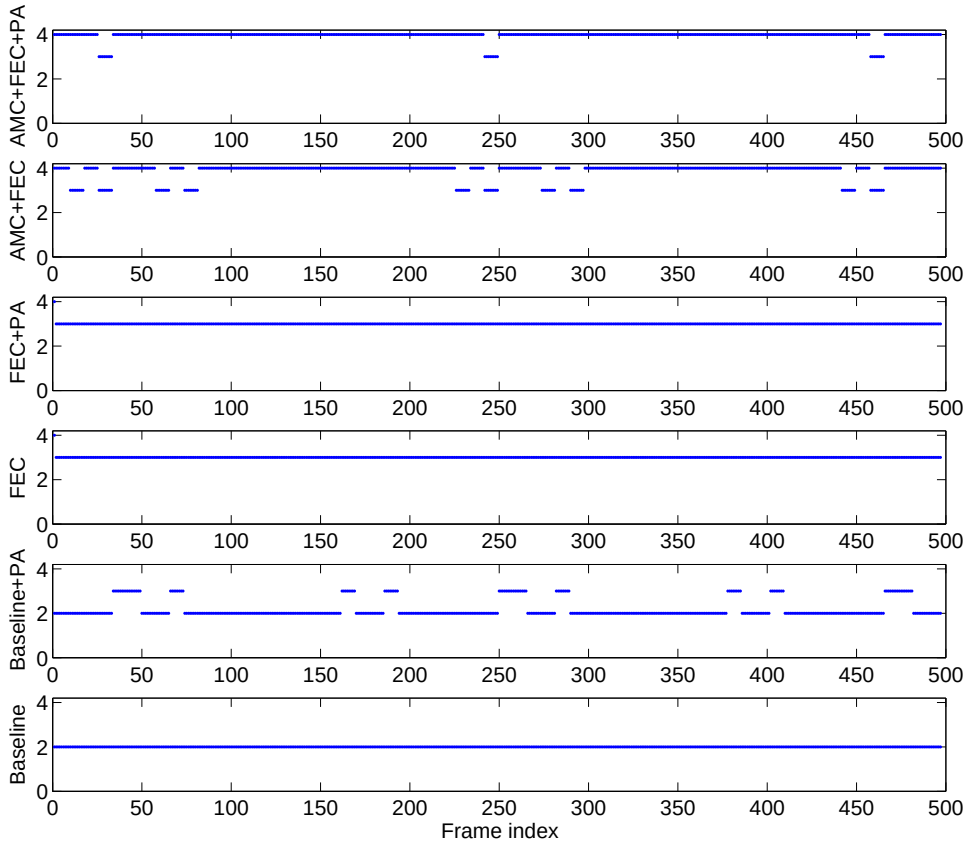

Figure 4.6: SVC layer indices of received frames. "Cactus", SNR: 10 dB.

channel gain. In this case, to meet the bit rate requirement of each SVC layer, the MCS of transmitting the SVC layer 1 (base layer), 2, 3 and 4 are set as: QPSK 1/2, QPSK 1/2, 16-QAM 9/16 and 64-QAM 3/4 respectively. Also, the same MCSs are applied for all the schemes without AMC for fair comparisons. The "Baseline+PA" scheme is similar to the work in [10], where the power allocation algorithm is applied on the "Baseline" system.

Figure 4.6 shows the received SVC layer index of each decoded video frame of "Cactus" when the system SNR is 10 dB. It is obvious that the proposed packet mapping and FEC rate adaptation scheme (i.e., "FEC" scheme) has more video frames with higher enhancement layers decoded. This is because the proposed scheme can better exploit the multi-channel

diversity gain. If the AMC scheme is applied (i.e., "AMC+FEC" scheme), the performance of the proposed scheme can be further improved since the best MCSs are chosen according to different channel qualities. The best system performance is achieved when both AMC and power allocation are applied with FEC rate adaptation (i.e., "AMC+FEC+PA" scheme).

The cumulative distribution functions (CDF) of the power allocated on each spatial channel are plotted in Fig. 4.7. The "Baseline+PA" scheme allocates power based on the fact that each SVC video layer is transmitted through a single spatial channel. While both "AMC+FEC+PA" and "FEC+PA" schemes allocates power based on the algorithm in Table 4.3, which minimizes the weighted sum of packet loss rate. Since the "AMC+FEC+PA" scheme adoptes different MCSs in each coherence time, its power allocation result is different to the "FEC+PA" scheme.

The per-frame MS-SSIM and PSNR of decoded video "Cactus" at system SNR 10 dB are plotted in Fig. 4.8 and Fig. 4.9 respectively. The effectiveness of the proposed algorithms are clearly demonstrated by comparing with "Baseline+PA" and "Baseline" methods.

The average power and the average spectral efficiency on each spatial channel under different channel conditions are illustrated in Fig. 4.10 and Fig. 4.11 respectively. Note that the sepctral efficiency depends on both the MCSs and the spatial channel selections. If a specific spatial channel is not selected, its spectral efficiency is 0. Also, the spectral efficiency can not reflect the successful decoding rate at the APP layer. For instance, consider the "Baseline" scheme, the spectral efficiencies of the 3^{rd} and 4^{th} channel are higher than the other schemes. However, due to lack of protection, the transmitted packets of these two channels cannot be successfully decoded. Figure 4.12 shows the average APP-FEC coding rates of each SVC video layer.

The decoded videos of "Cactus" of "AMC+FEC+PA", "FEC" and "Baseline" schemes at system SNR 10 dB are available for reviews at [1]. One can observe that the video decoding quality is the best with "AMC+FEC+PA" method.

Figure 4.13 shows the average MS-SSIM index and the average PSNR at different average channel SNRs. Apparently, without AMC and power allocation, the proposed scheme (i.e., "FEC" scheme) still outperforms "Baseline" and "Baseline+PA" schemes. However, with fixed MCSs, the system throughput is limited and the system performance cannot be

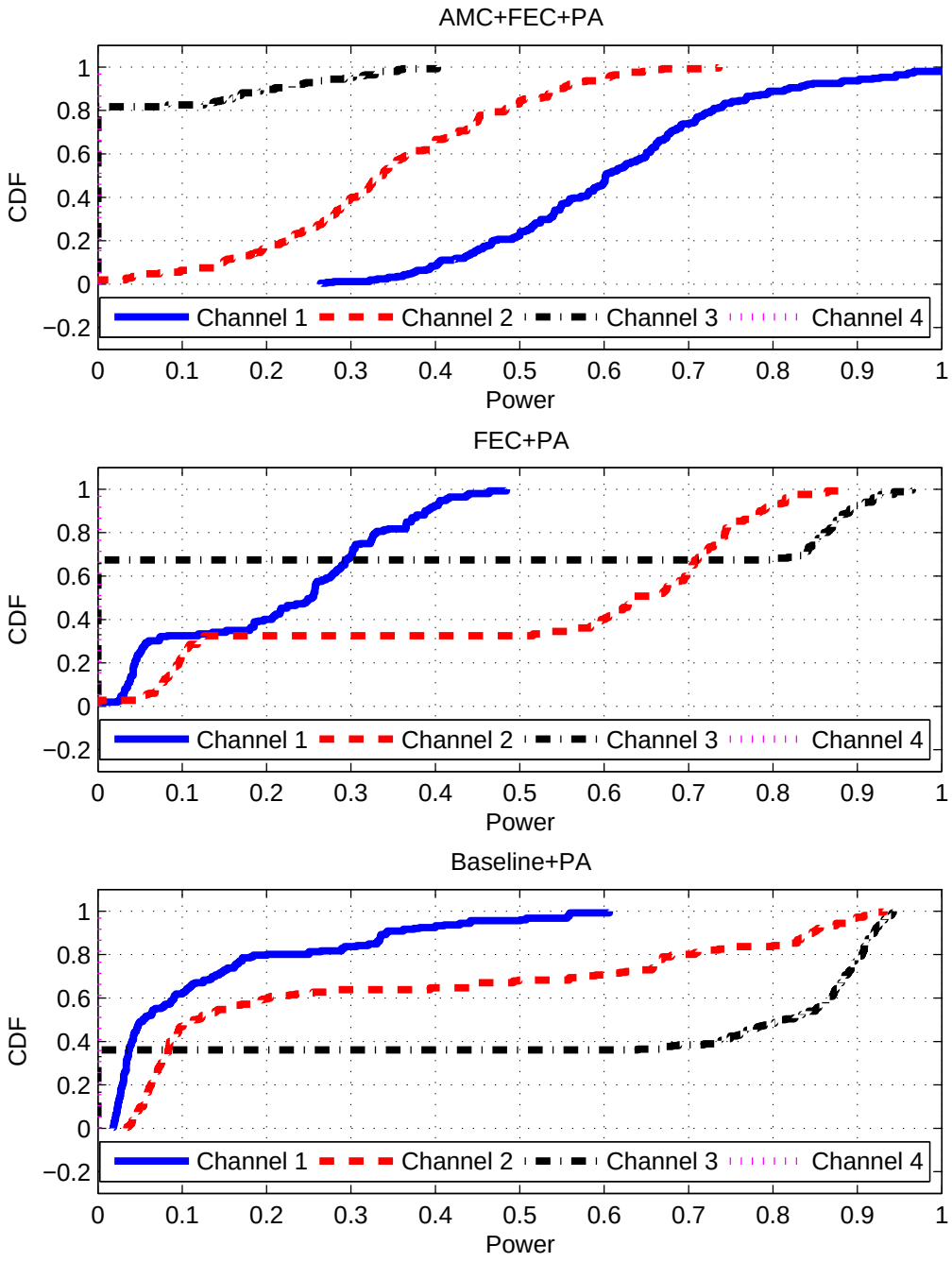

Figure 4.7: CDF of power on each spatial channel. "Cactus", SNR: 10 dB.

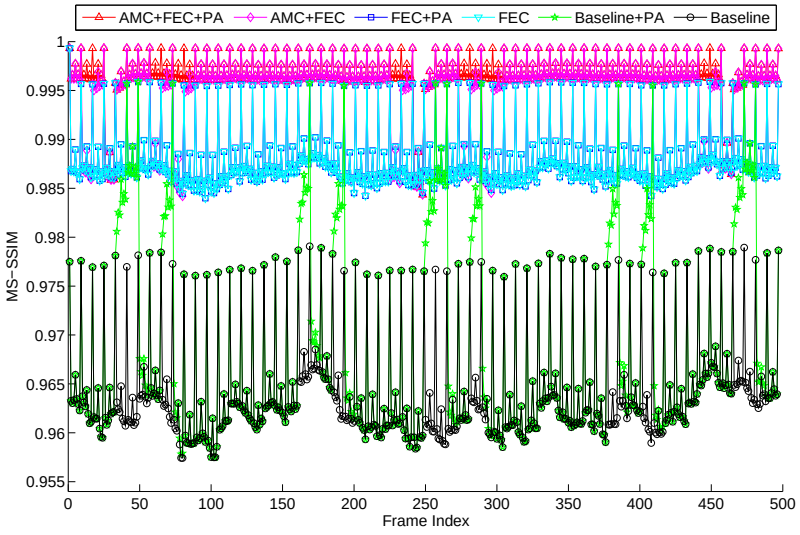

Figure 4.8: Per-frame MS-SSIM indices of video. "Cactus", SNR: 10 dB.

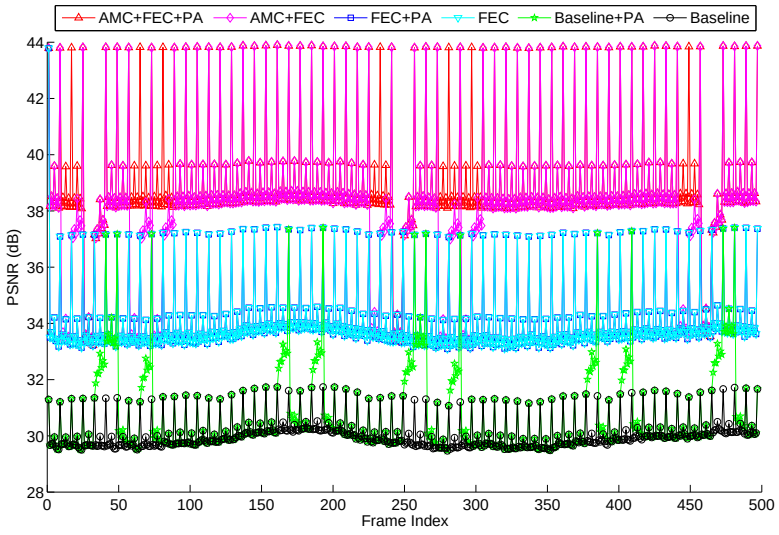

Figure 4.9: Per-frame PSNR indices of video. "Cactus", SNR: 10 dB.

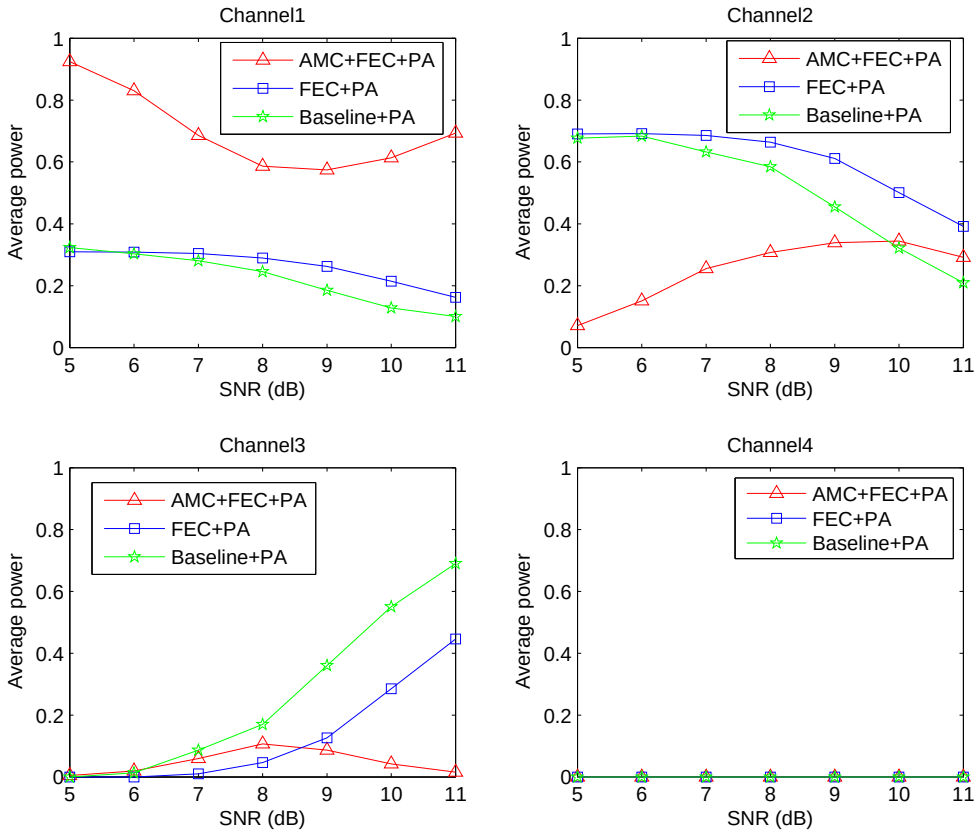

Figure 4.10: Average power allocated on each spatial channel. “Cactus”.

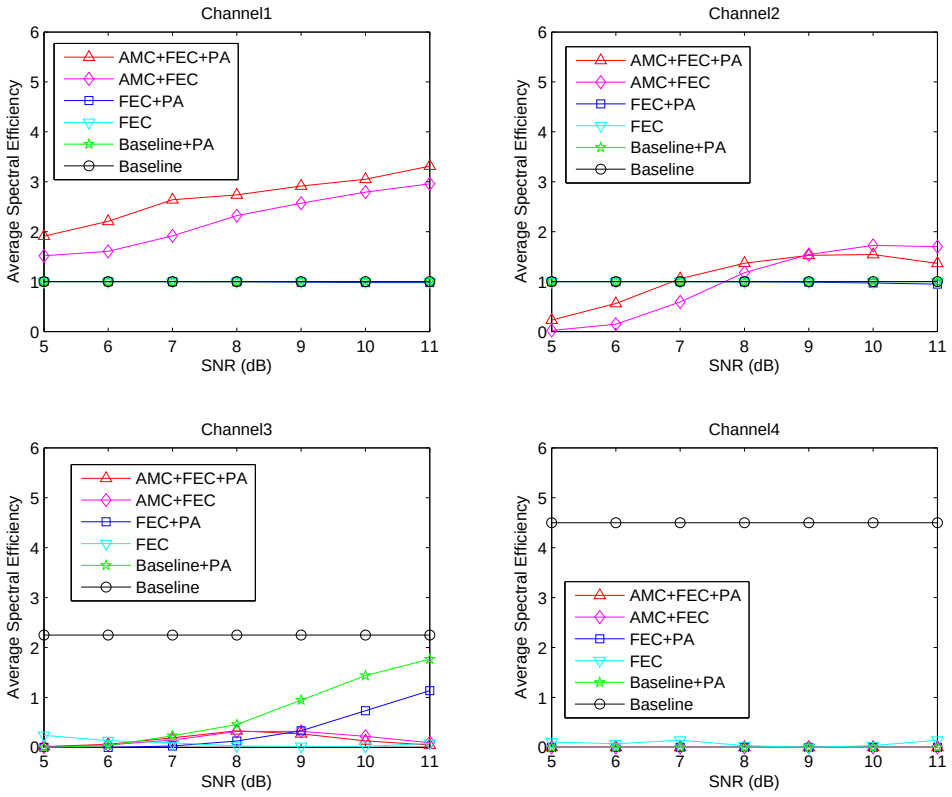

Figure 4.11: Average spectral efficiency on each spatial channel. “Cactus”.

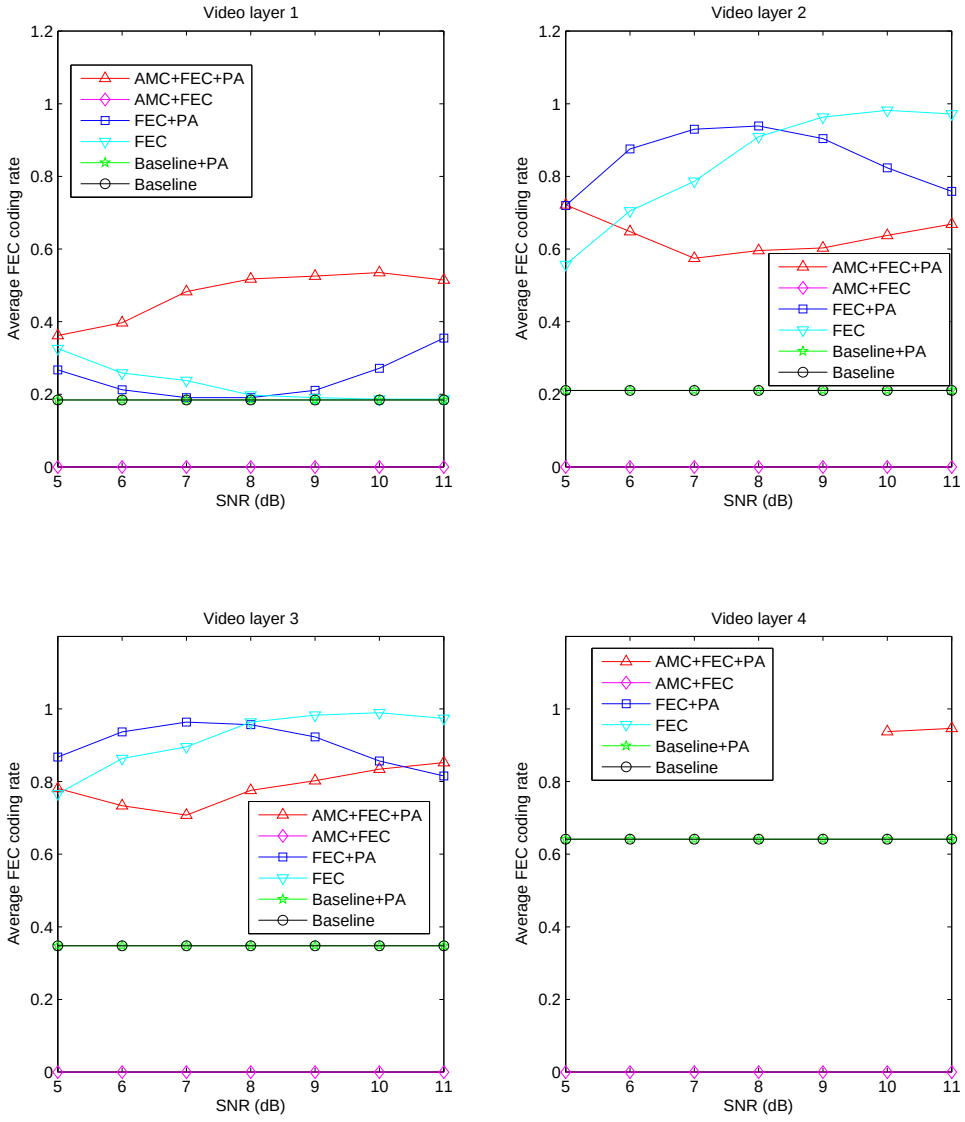

Figure 4.12: Average APP-FEC coding rate of each video layer. “Cactus”.

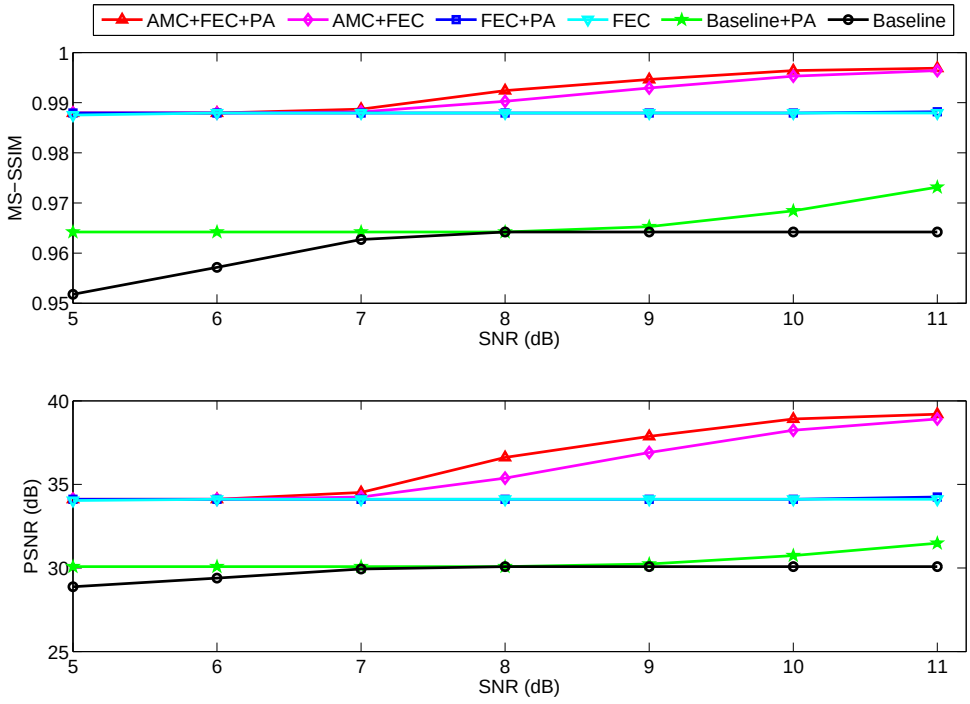

Figure 4.13: Average MS-SSIM index and PSNR of reconstructed video “Cactus”.

further improved even when channel quality becomes better. If the AMC scheme is applied (i.e., "AMC+FEC" scheme), decoded video quality becomes better when channel quality improves. Power allocation can provide additional decoding quality gain as shown in the "AMC+FEC+PA" scheme.

Similar results can be demonstrated by the video clip "Kimono". Figure 4.14 shows the SVC quality layer indices of the reconstructed video when system SNR is 12 dB. The CDFs of the power allocated on each spatial channel is shown in Fig. 4.15. The corresponding MS-SSIM and PSNR plots are shown in Fig. 4.16 and Fig. 4.17 respectively. Clearly, the proposed methods outperform the others. The average power and spectral efficiency of each spatial channel are plotted in Fig. 4.18 and Fig. 4.19 respectively. The average APP-FEC coding rates of each SVC layer are plotted in Fig. 4.20. The visual comparison of the decoded frame of "AMC+FEC+PA", "FEC" and "Baseline" methods at system SNR 12 dB are available at [1]. Since AMC+FEC+PA can decoded more enhancement layers than the others, its video quality is the best. The average MS-SSIM indices and average PSNR at different system SNR are plotted in Fig. 4.21. Still, the proposed rate adaptation and power allocation scheme (i.e., "AMC+FEC+PA") has the best performance among the other schemes.

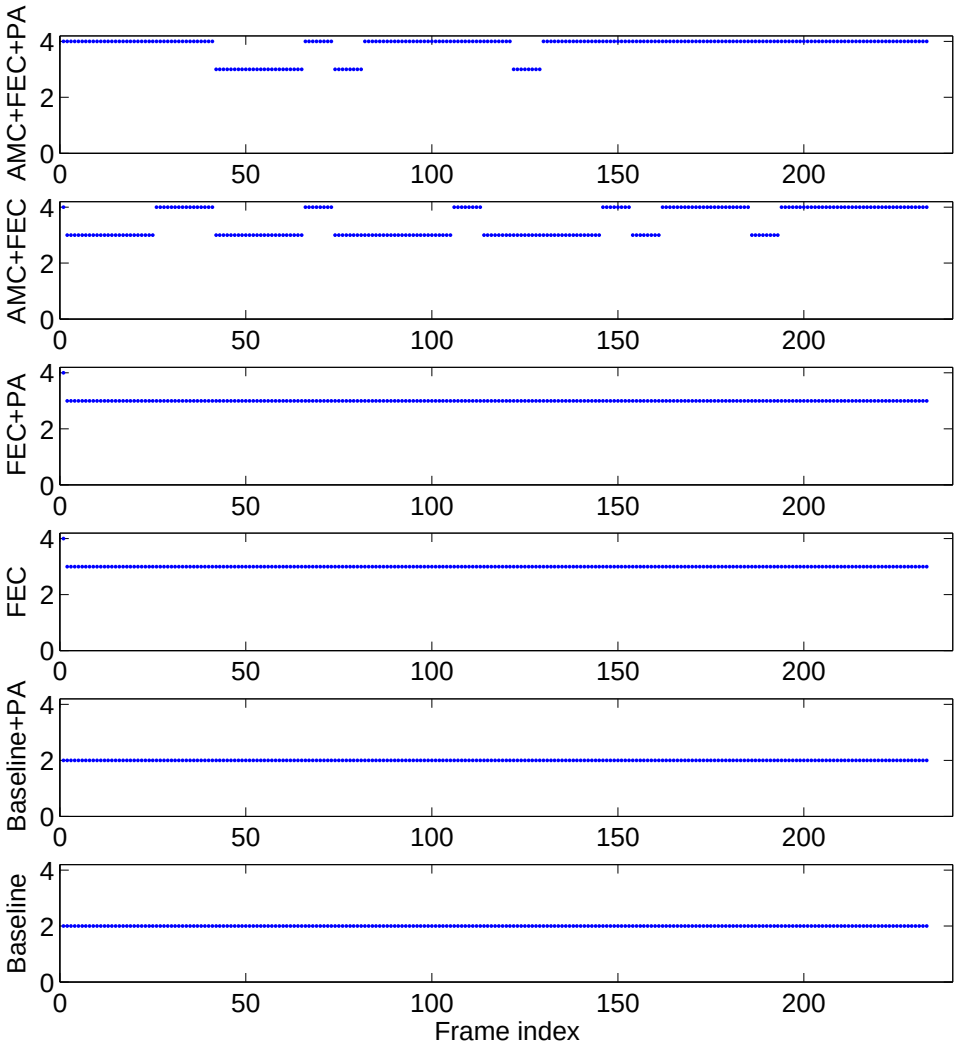

Figure 4.14: SVC layer indices of received frames. “Kimono”, SNR: 12 dB.

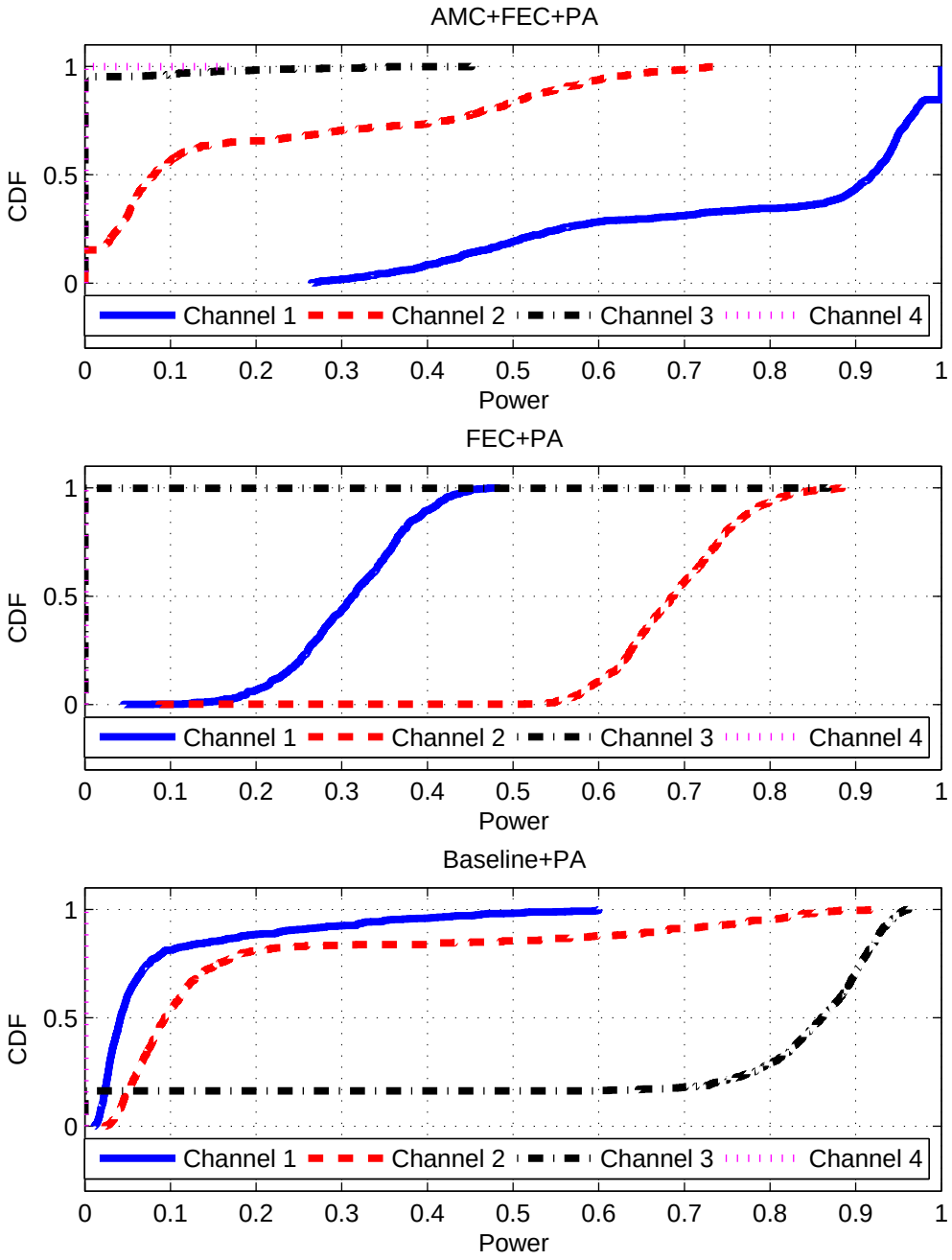

Figure 4.15: CDF of power on each spatial channel. "Kimono", SNR: 12 dB.

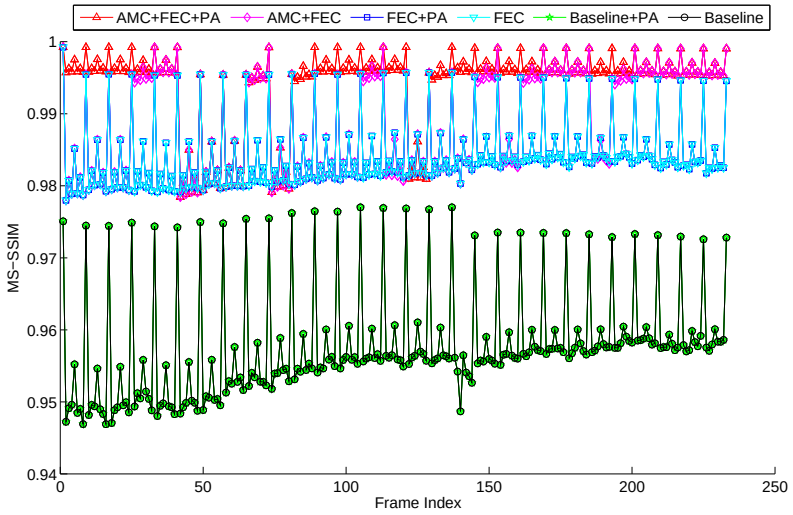

Figure 4.16: Per-frame MS-SSIM indices of video. "Kimono", SNR: 12 dB.

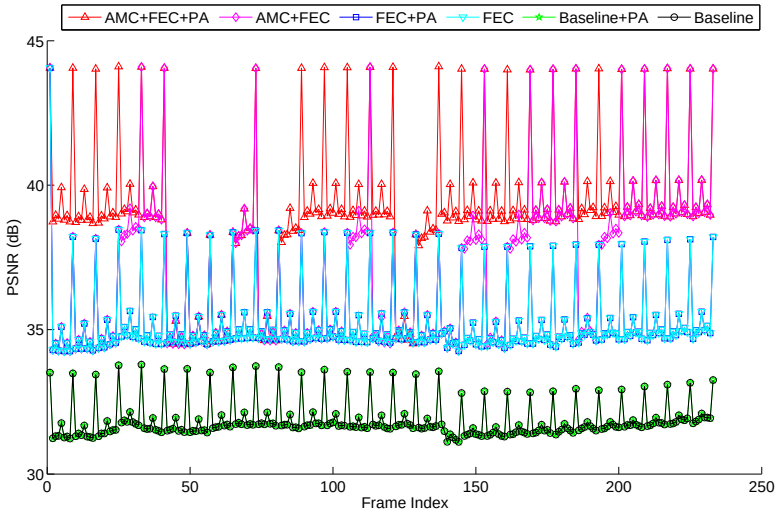

Figure 4.17: Per-frame PSNR indices of video. "Kimono", SNR: 12 dB.

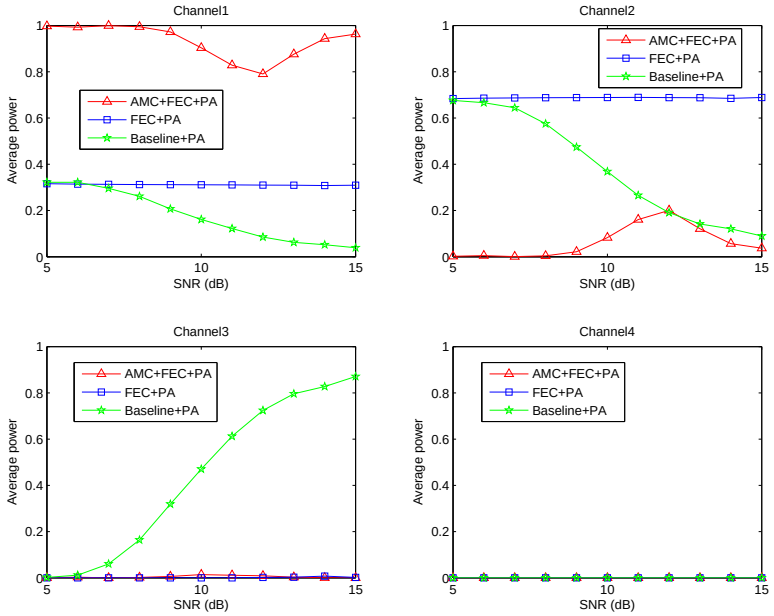

Figure 4.18: Average power allocated on each spatial channel. "Kimono".

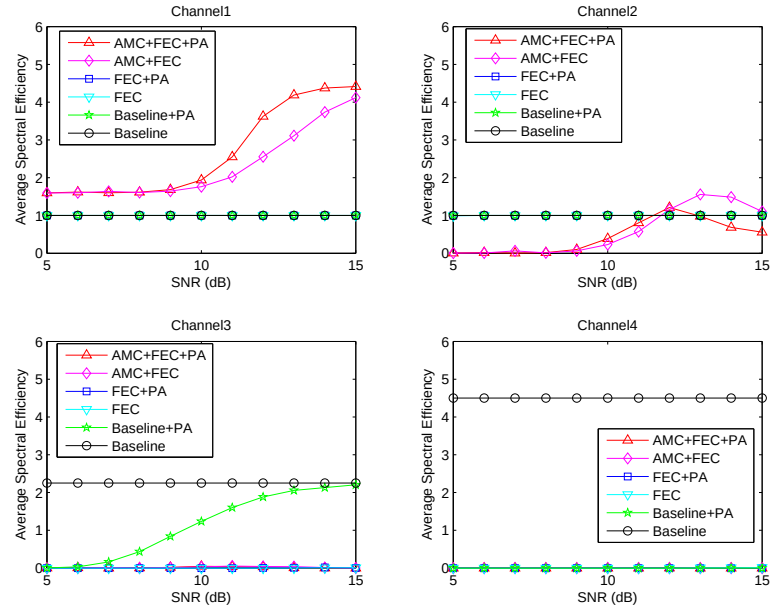

Figure 4.19: Average spectral efficiency on each spatial channel. "Kimono".

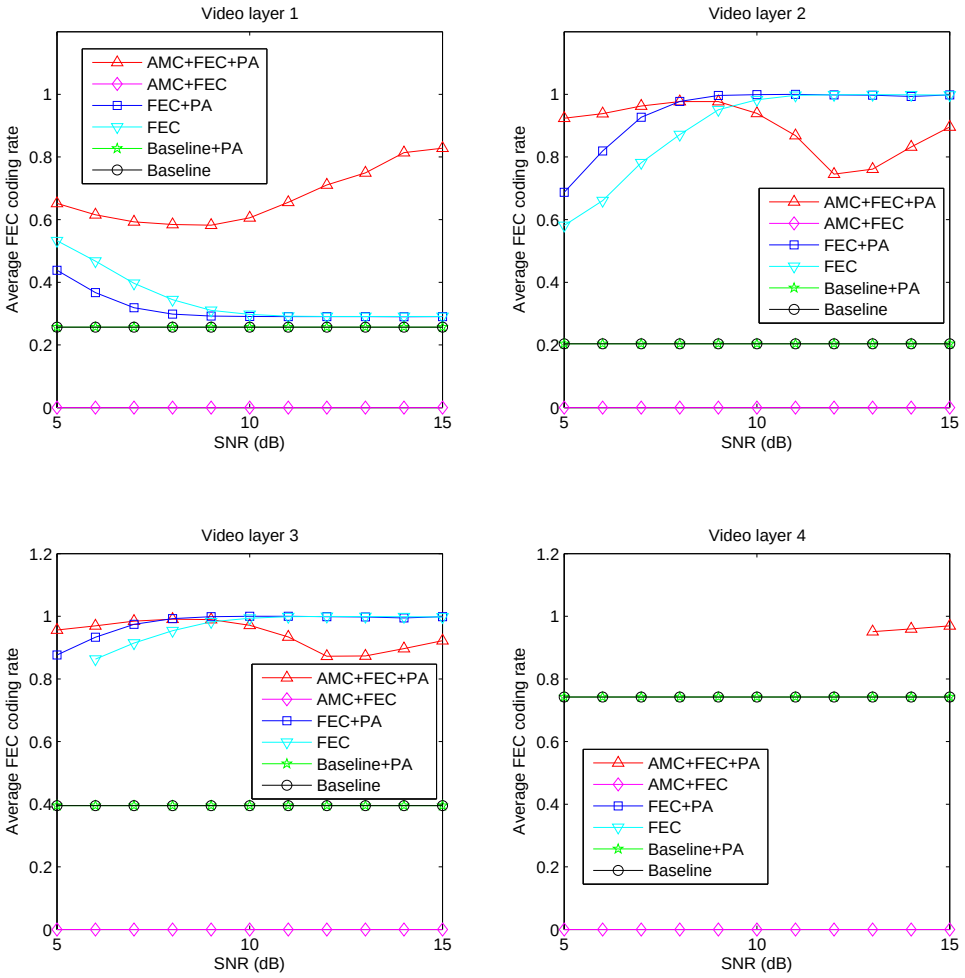

Figure 4.20: Average APP-FEC coding rate of each video layer. “Kimono”.

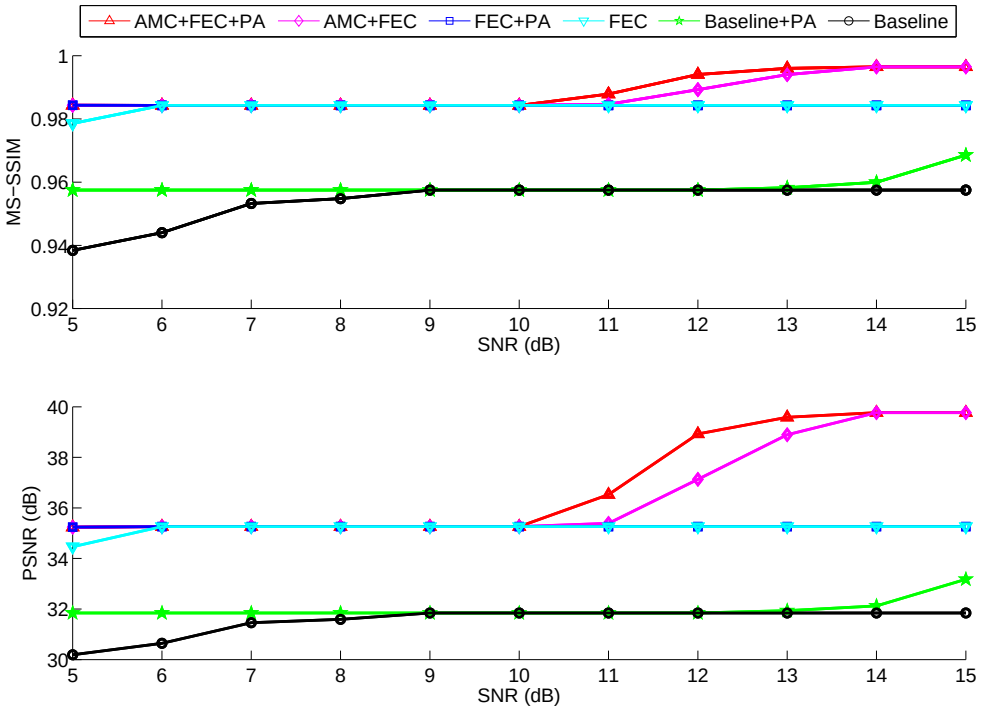

Figure 4.21: Average MS-SSIM index and PSNR of reconstructed video "Kimono".

Chapter 5

CONCLUSIONS AND FUTURE WORK

In this thesis, the SVC-based video transmission schemes over MIMO systems are discussed and evaluated. First, a near-optimal transmit power allocation scheme is discussed. By allocating appropriate power on each spatial channel, the SVC layers are transmitted by UEPs, which can achieve higher QoE at the end users. To reduce the complexity of this problem, the original problem is decomposed into several convex sub-problems and the optimal solution can be solved by a bisection search algorithm. A low-complexity algorithm is also developed to further reduce the complexity with an acceptable performance loss. Second, a joint rate and power adaptation scheme is discussed. In this work, the bit stream of each SVC layer is spread to multiple MIMO spatial channels. By applying APP-FEC, the spatial diversity gain can be exploited, which can significantly improve the decoding qualities at the end users. With proper source and channel coding rate adaptations, each SVC layer is transmitted with UEP so that more important bit streams are better protected. Transmit power allocation is applied to further improve the system performance. The original joint optimization problem can be decomposed into several sub-problems. Each sub-problem can be solved by series of well-developed algorithms so that the overall complexity can be reduced. The effectiveness and favorable performance of the two proposed schemes are evaluated by plenty of simulations with H.264 SVC traces of CIF and HD video clips.

Many topics can be studied in the future. First, orthogonal frequency-division multiplexing (OFDM) is one of the key technologies in modern wireless communication systems due to its high bandwidth efficiency and ability of combating frequency selective fading channels [61]. With OFDM, the broadband frequency selective fading channels can be decomposed into several parallel narrowband frequency flat fading channels with different channel gains [19]. Thus, OFDM enables another dimension (frequency domain) for wireless

data transmissions. Power allocation on frequency domain can be applied to increase the theoretical capacity through the well-known WF algorithm [59]. Therefore, similar quality-driven rate and power adaptation scheme can be studied if MIMO-OFDM is considered. Second, the proposed two schemes in this thesis are based on the assumption of full and accurate channel knowledge at both transmitter and receiver sides. However, in reality, only partial and inaccurate channel knowledge is available. Therefore, a more robust system is required to be developed so that the video decoding quality can still be acceptable if the transmission system is lack of full accurate channel knowledge. Third, in this thesis, the optimiaztion target of the proposed schemes are better video decoding qualities at the end users. Nevertheless, there are plenty of other applications in video transmissions. For instance, videos can be transmitted to cloud server for video analytics such as human/car detections and tracking, behavior understandings and action recognitions [21, 28, 37–42], especially for video surveillance systems [78] with the emerging of the Internet of Things (IoT). Therefore, instead of transmitting videos for human perception purposes, novel cross-layer video transmission schemes can be studied so that the videos are delivered for better video analytics performance [12, 13].

BIBLIOGRAPHY

[1] The comparisons of decoded videos [Online]. Available at http://allison.ee.washington.edu/xchen/svc_mimo/.

[2] The JSVM (joint scalable video model) software [Online]. Available at http://www.hhi.fraunhofer.de/de/kompetenzfelder/image-processing/research-groups/image-video-coding/svc-extension-of-h264avc/jsvm-reference-software.html.

[3] "IEEE Standard for Local and Metropolitan Area Networks Part 16: Air Interface for Broadband Wireless Access Systems," *IEEE Std. 802.16-2009*, 2009.

[4] "Cisco Visual Networking Index: Forecast and Methodology, 2014-2019," 2015.

[5] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[6] S. Catreux, V. Erceg, D. Gesbert, and R. W. Heath Jr, "Adaptive modulation and mimo coding for broadband wireless data networks," *IEEE Communications Magazine*, vol. 40, no. 6, pp. 108–115, 2002.

[7] C.-H. Chen, W.-H. Chung, and Y.-C. F. Wang, "Cross-layer design for video streaming with dynamic antenna selection," in *Proc. of IEEE Intl. Conf. on Image Processing*, Brussels, September 11-14 2011, pp. 3245–3248.

[8] X. Chen, "Simulation of superposition coding," 2009.

[9] X. Chen, H. Du, J.-N. Hwang, J. A. Ritcey, and C.-N. Lee, "A QoE-driven FEC rate adaptation scheme for scalable video transmissions over MIMO systems," in *Proc. of IEEE Intl. Conf. on Communications*, London, UK, June 8-12 2015, pp. 6953–6958.

[10] X. Chen, J.-N. Hwang, C.-N. Lee, and S.-I. Chen, "A near optimal QoE-driven power allocation scheme for scalable video transmissions over MIMO systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 1, pp. 76–88, 2015.

[11] X. Chen, J.-N. Hwang, C.-N. Lee, and C.-W. Hwang, "An efficient CQI feedback resource allocation scheme for wireless video multicast services," in *Proc. of IEEE Global Communications Conf.*, Atlanta, GA, December 9-13 2013, pp. 1663–1668.

[12] X. Chen, J.-N. Hwang, K.-H. Lee, and R. L. de Queiroz, "Quality-of-content (QoC)-driven rate allocation for video analysis in mobile surveillance networks," in *Proc. of IEEE Intl. Workshop on Multimedia Signal Processing*, Xiamen, China, October 19-21 2015.

[13] X. Chen, J.-N. Hwang, D. Meng, K.-H. Lee, R. L. de Queiroz, and F.-M. Yeh, "Quality-of-content (QoC)-based joint source and channel coding for human detections in a mobile surveillance cloud."

[14] X. Chen, J.-N. Hwang, J. A. Ritcey, C.-N. Lee, and F.-M. Yeh, "Quality-driven joint rate and power adaptation for scalable video transmissions over MIMO systems," *IEEE Trans. on Circuits and Systems for Video Technology*, 2016.

[15] X. Chen, J.-N. Hwang, H.-J. Su, and C.-N. Lee, "Adaptive mode and modulation coding switching scheme in MIMO multicasting system," in *Proc. of IEEE Intl. Symp. on Circuits and Systems*, Beijing, China, May 19-23 2013.

[16] X. Chen, J.-N. Hwang, H. J. Su, and C. N. Lee, "System for selecting transmission mode under multi-input multi-output based on scheduling number and method thereof," Patent US 20 150 016 556, 2013.

[17] X. Chen, J.-N. Hwang, C.-Y. Wang, and C.-N. Lee, "A near optimal QoE-driven power allocation scheme for SVC-based video transmissions over MIMO systems," in *Proc. of IEEE Intl. Conf. on Communications*, Sydney, Australia, June 10-14 2014.

[18] X. Chen, J.-N. Hwang, C.-J. Wu, S.-R. Yang, and C.-N. Lee, "A QoE-based APP layer scheduling scheme for scalable video transmissions over multi-RAT systems," in *Proc. of IEEE Intl. Conf. on Communications*, London, UK, June 8-12 2015.

[19] X. Chen, F. Qu, and L. Yang, "OFDM-IDMA for power line communications," in *Proc. of IEEE Intl. Symp. on Power Line Communications*, Udine, Italy, April 3-6 2011.

[20] S. Chikkerur, V. Sundaram, M. Reisslein, and L. J. Karam, "Objective video quality assessment methods: A classification, review, and performance comparison," *IEEE Trans. on Broadcasting*, vol. 57, no. 2, pp. 165–182, 2011.

[21] C.-T. Chu, K.-H. Lee, and J.-N. Hwang, "Self-organized and scalable camera networks for systematic human tracking across nonoverlapping cameras," in *Proc. of IEEE Intl. Conf. on Acoustics, Speech, and Signal Processing*, Vancouver, BC, May 26-31 2013, pp. 2322–2326.

[22] S.-P. Chuah, Y.-P. Tan, and Z. Chen, "Rate and power allocation for joint coding and transmission in wireless video chat applications," *IEEE Trans. on Multimedia*, vol. 17, no. 5, pp. 687–699, 2015.

[23] G. Dimić and N. D. Sidiropoulos, "On downlink beamforming with greedy user selection: performance analysis and a simple new algorithm," *IEEE Trans. on Signal Processing*, vol. 53, no. 10, pp. 3857–3868, 2005.

[24] M. Fiedler, T. Hossfeld, and P. Tran-Gia, "A generic quantitative relationship between quality of experience and quality of service," *IEEE Network*, vol. 24, no. 2, pp. 36–41, 2010.

[25] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs Technical Journal*, vol. 1, no. 2, pp. 41–59, 1996.

[26] W. Hamidouche, C. Olivier, Y. Pousset, and C. Perrine, "Optimal resource allocation for medium grain scalable video transmission over MIMO channels," *Journal of Visual Communication and Image Representation*, vol. 24, no. 3, pp. 373–387, 2013.

[27] W. Hamidouche, C. Perrine, Y. Pousset, and C. Olivier, "A solution to efficient power allocation for H.264/SVC video transmission over a realistic MIMO channel using precoder designs," *Journal of Visual Communication and Image Representation*, vol. 22, no. 6, pp. 563–574, 2011.

[28] L. Hou, W. Wan, K.-H. Lee, and J.-N. Hwang, "Deformable multiple-kernel based human tracking using a moving camera," in *Proc. of IEEE Intl. Conf. on Acoustics, Speech, and Signal Processing*, South Brisbane, QLD, April 19-24 2015.

[29] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?" *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 2, pp. 160–171, 2013.

[30] C.-W. Huang, S.-M. Huang, P.-H. Wu, S.-J. Lin, and J.-N. Hwang, "OLM: Opportunistic layered multicasting for scalable IPTV over mobile WiMAX," *IEEE Trans. on Mobile Computing*, vol. 11, no. 3, pp. 453–463, 2012.

[31] K. Huang, J. G. Andrews, and J. Robert W. Heath, "Performance of orthogonal beamforming for SDMA with limited feedback," *IEEE Trans. on Vehicular Tech.*, vol. 58, no. 1, pp. 152–164, 2009.

[32] J.-N. Hwang, *Multimedia Networking: From Theory to Practice*. Cambridge University Press, 2009.

[33] M. K. Jubran, M. Bansal, L. P. Kondi, and R. Grover, "Accurate distortion estimation and optimal bandwidth allocation for scalable H.264 video transmission over MIMO systems," *IEEE Trans. on Image Processing*, vol. 18, no. 1, pp. 106–116, 2009.

[34] D. Jurca, P. Frossard, and A. Jovanovic, "Forward error correction for multipath media streaming," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 19, no. 9, pp. 1315–1326, 2009.

[35] E. Karipidis, N. D. Sidiropoulos, and Z.-Q. Luo, "Quality of service and max-min fair transmit beamforming to multiple cochannel multicast groups," *IEEE Trans. on Signal Processing*, vol. 56, no. 3, pp. 1268–1279, 2008.

[36] H. Kim, P. C. Cosman, and L. B. Milstein, "Motion-compensated scalable video transmission over mimo wireless channels," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 23, no. 1, pp. 116–127, 2013.

[37] K.-H. Lee and J.-N. Hwang, "On-road pedestrian tracking across multiple driving recorders," *IEEE Trans. on Multimedia*, vol. 17, no. 9, pp. 1429–1438, 2015.

[38] K.-H. Lee, J.-N. Hwang, and S.-I. Chen, "Model-based vehicle localization based on three-dimensional constrained multiple-kernel tracking," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 25, no. 1, pp. 38–50, 2014.

[39] K.-H. Lee, J.-N. Hwang, G. Okapal, and J. Pitton, "Driving recorder based on-road pedestrian tracking using visual SLAM and constrained multiple-kernel," in *Proc. of IEEE Intl. Conf. Intelligent Transportation Systems*, Qingdao, China, October 8-11 2014.

[40] K.-H. Lee, J.-N. Hwang, J.-H. Yoo, and K.-H. Choi, "Effective car video retrieval using feature matching in a mobile video cloud," in *Proc. of ACM/IEEE Intl. Conf. Distributed Smart Cameras*, Hong Kong, October 30- November 2 2012.

[41] K.-H. Lee, J.-N. Hwang, J.-Y. Yu, and K.-Z. Lee, "Vehicle tracking iterative by kalman-based constrained multiple-kernel and 3-D model-based localization," in *Proc. of IEEE Intl. Symp. on Circuits and Systems*, Beijing, China, May 19-23 2013.

[42] K.-H. Lee, Y.-J. Lee, and J.-N. Hwang, "Multiple-kernel based vehicle tracking using 3-D deformable model and license plate self-similarity," in *Proc. of IEEE Intl. Conf. on Acoustics, Speech, and Signal Processing*, Vancouver, BC, May 26-31 2013.

[43] J. Li, Z. Bao, C. Zhang, and Q. Li, "Scalable video multicast with joint resource allocation and adaptive modulation and coding over multiple base station networks," in *Proc. of IEEE Intl. Conf. on Networking, Sensing and Control (ICNSC)*, 2015, pp. 134–139.

[44] M. Li, Z. Chen, and Y.-P. Tan, "Scalable resource allocation for svc video streaming over multiuser mimo-ofdm networks," *IEEE Trans. on Multimedia*, vol. 15, no. 7, pp. 1519–1531, 2013.

[45] Q. Li, G. Li, W. Lee, M. il Lee, D. Mazzarese, B. Clerckx, and Z. Li, "MIMO techniques in WiMAX and LTE: a feature overview," *IEEE Communications Magazine*, vol. 48, no. 5, pp. 86–92, 2010.

[46] L. Liu, R. Chen, S. Geirhofer, K. Sayana, Z. Shi, and Y. Zhou, "Downlink mimo in LTE-advanced: SU-MIMO vs. MU-MIMO," *IEEE Communications Magazine*, vol. 50, no. 2, pp. 140–147, 2012.

[47] Q. Liu, S. Liu, and C. Chen, "A novel prioritized spatial multiplexing for MIMO wireless system with application to H.264 SVC video," in *Proc. of IEEE Intl. Conf. on Multimedia and Expo*, Suntec City, July 19-23 2010.

[48] M. Luby, T. Stockhammer, and M. Watson, "Application layer FEC in IPTV services," *IEEE Communications Magazine*, vol. 46, no. 5, pp. 94–101, 2008.

[49] H. Luo, D. Wu, S. Ci, H. Sharif, and H. Tang, "TFRC-based rate control for real-time video streaming over wireless multi-hop mesh networks," in *Proc. of IEEE Intl. Conf. on Communications*, Dresden, June 14-18 2009.

[50] Z. Ma, M. Xu, Y.-F. Ou, and Y. Wang, "Modeling of rate and perceptual quality of compressed video as functions of frame rate and quantization stepsize and its applications," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 22, no. 5, pp. 671–682, 2012.

[51] A. K. Moorthy, K. Seshadrinathan, R. Soundararajan, and A. C. Bovik, "Wireless video quality assessment: A study of subjective scores and objective algorithms," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 20, no. 4, pp. 587–599, 2010.

[52] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Trans. on Communications*, vol. 61, no. 4, pp. 1436–1449, 2013.

[53] C. Oestges and B. Clerckx, *MIMO Wireless Communications From Real-World Propagation to Space-Time Code Design.* Academic Press, 2007.

[54] J. Park, X. Chen, and J.-N. Hwang, "Optimum power allocation and rate adaptation for scalable video streaming over multi-user MIMO networks," in *Proc. of IEEE Global Communications Conf.*, San Diego, CA, December 6-10 2015.

[55] A. Paulraj, R. Nabar, and D. Gore, *Introduction to Space-Time Wireless Communications.* Cambridge Press, 2003.

[56] A. J. Paulraj, D. A. Gore, R. U. Nabar, and H. Bölcskei, "An overview of MIMO communicationsa key to gigabit wireless," *Proceedings of the IEEE*, vol. 92, no. 2, pp. 198–218, 2004.

[57] A. J. Paulraj and T. Kailath, "Increasing capacity in wireless broadcast systems using distributed transmission/directional reception," Patent US 5 345 599, 1994.

[58] C. Perkins, *RTP Audio and Video for the Internet.* Addison Wesley, 2003.

[59] J. G. Proakis and M. Salehi, *Digital Communications.* McGraw Hill, 2008.

[60] F. Qu, X. Chen, P. Deng, and L. Yang, "Iterative MLSE for MIMO underwater acoustic channel," in *Proc. of MTS/IEEE Oceans Conference*, Seattle, WA, September 20-23 2010.

[61] F. Rey, M. Lamarca, and G. Vazquez, "Robust power allocation algorithms for MIMO-OFDM systems with imperfect CSI," pp. 1070–1085, 2005.

[62] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Processing Magazine*, vol. 30, no. 1, pp. 40–60, 2013.

[63] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable video coding extension of the H.264/AVC standard," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 17, no. 9, pp. 1103–1120, 2007.

[64] S. Sesia, I. Toufik, and M. Baker, *LTE: the UMTS long term evolution.* Wiley Online Library, 2009.

[65] A. Shokrollahi, "Raptor codes," *IEEE Trans. on Information Theory*, vol. 52, no. 6, pp. 2551–2567, 2006.

[66] N. D. Sidiropoulos, T. N. Davidson, and Z.-Q. Luo, "Transmit beamforming for physical-layer multicasting," *IEEE Trans. on Signal Processing*, vol. 54, no. 6, pp. 2239–2251, 2006.

[67] Y. C. B. Silva and A. Klein, "Linear transmit beamforming techniques for the multi-group multicast scenario," *IEEE Trans. on Vehicular Tech.*, vol. 58, no. 8, pp. 4353–4367, 2009.

[68] B. Sklar, *Digital Communication - Fundamentals and Applications.* Prentice Hall, 2001.

[69] D. Song and C. Chen, "Scalable H.264/AVC video transmission over MIMO wireless systems with adaptive channel selection based on partial channel information," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 17, no. 9, pp. 1218–1226, 2007.

[70] ——, "Maximum-throughput delivery of SVC-based video over MIMO systems with time-varying channel capacity," *Journal of Visual Communication and Image Representation*, vol. 19, no. 8, pp. 520–528, 2008.

[71] H. Stark and J. Woods, *Probability, Statistics, and Random Processes for Engineers*. Prentice Hall, 2011.

[72] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: Performance criterion and code construction," *IEEE Trans. on Information Theory*, vol. 44, no. 2, pp. 744–765, 1998.

[73] M. Vranješ, S. Rimac-Drlje, and K. Grgić, "Review of objective video quality metrics and performance comparison using different databases," *Signal Processing: Image Communication*, vol. 28, no. 1, pp. 1–19, 2013.

[74] S.-C. Wang and W. Liao, "Cooperative multicasting for wireless scalable video transmissions," *IEEE Trans. on Communications*, vol. 61, no. 9, pp. 3980–3989, 2013.

[75] Z. Wang, L. Lu, and A. C. Bovik, "Video quality assessment based on structural distortion measurement," *Signal Processing: Image Communication*, vol. 19, no. 2, pp. 121–132, 2004.

[76] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multi-scale structural similarity for image quality assessment," in *Proc. of IEEE Asilomar Conf. Signals, Systems and Computers*, November 9-12 2003.

[77] T. Wiegand, L. Noblet, and F. Rovati, "Scalable video coding for IPTV services," *IEEE Trans. on Broadcasting*, vol. 55, no. 2, pp. 527–538, 2009.

[78] P.-H. Wu, C.-W. Huang, J.-N. Hwang, J. young Pyun, and J. Zhang, "Video-quality-driven resource allocation for real-time surveillance video uplinking over OFDMA-based wireless networks," *IEEE Trans. on Vehicular Tech.*, vol. 64, no. 7, pp. 3233–3246, 2014.

[79] J. Xu, R. Hormis, and X. Wang, "Mimo video broadcast via transmit-precoding and snr-scalable video coding," *IEEE Journal on Selected Areas in Communications*, vol. 28, no. 3, pp. 456–466, 2010.

[80] Y. Yang, "Contributions to smart metering protocol design and data analytics," Ph.D. dissertation, University of Washington, 2015.

[81] Y. Yang and S. Roy, "PMU deployment for optimal state estimation performance," in *Proc. of IEEE Global Communications Conf.*, Anaheim, CA, December 3-7 2012.

[82] ——, "PMU deployment for three-phase optimal state estimation performan," in *Proc. of IEEE Intl. Conf. on Smart Grid Communications*, Vancouver, BC, October 21-24 2013.

[83] ——, "Grouping based MAC protocols for EV charging data transmission in smart metering network," *IEEE Journal on Selected Areas in Communications*, vol. 49, no. 7, pp. 1328–1343, 2014.

[84] ——, "PCF scheme for periodic data transmission in smart metering network with cognitive radio," in *Proc. of IEEE Global Communications Conf.*, San Diego, CA, December 6-10 2015.

[85] Y. Yang, X. Wang, and X. Cai, "On the number of relays for orthogonalize-and-forward relaying," in *Proc. of IEEE Intl. Conf. on Wireless Communications and Signal Processing*, Nanjing, China, November 9-11 2011.

[86] Z. Yang and Y. Zhao, "Scalable video multicast over multi-antenna OFDM systems," *Wireless Personal Communications*, vol. 70, no. 4, pp. 1487–1504, 2013.

[87] T. Yoo and A. Goldsmith, "On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 3, pp. 528–541, 2006.

[88] L. Zheng and D. N. C. Tse, "Diversity and multiplexing: a fundamental tradeoff in multiple-antenna channels," *IEEE Trans. on Information Theory*, vol. 49, no. 5, pp. 1073–1096, 2003.

Appendix A

A.1 Derivation of Eq. (3.15) By Induction

Consider (3.10), if $t = 0$,$\mathrm{Pe}_l(p_l) = P_{M_l}(p_l)$. If $t = 1$,

$$\begin{aligned}\mathrm{Pe}_l(p_l) &= P_{M_l}(p_l) - \frac{1}{N} N\left(P_{M_l}(p_l)\right)\left(1 - P_{M_l}(p_l)\right)^{N-1} \\ &\approx (N-1)\, P_{M_l}(p_l)^2\end{aligned} \tag{A.1}$$

with first order Taylor series approximation if $P_{M_l}(p_l) \ll 1$. If $t = 2$,

$$\begin{aligned}\mathrm{Pe}_l(p_l) &\approx (N-1)\, P_{M_l}(p_l)^2 - \frac{2}{N}\binom{N}{2} P_{M_l}(p_l)^2 \left(1 - P_{M_l}(p_l)\right)^{N-2} \\ &\approx (N-1)(N-2)\, P_{M_l}(p_l)^3 .\end{aligned} \tag{A.2}$$

Similar steps can be followed when $t = 3$,

$$\mathrm{Pe}_l(p_l) \approx \frac{(N-1)! P_{M_l}(p_l)^3}{2\,(N-3)!} + \frac{(N-1)! P_{M_l}(p_l)^4}{2\,(N-4)!}. \tag{A.3}$$

When $t = 4$,

$$\mathrm{Pe}_l(p_l) \approx \sum_{j=3}^{4} \frac{(N-1)!\,(j-2)}{(N-j)!\,(j-1)!} P_{M_l}(p_l)^j + \frac{(N-1)! P_{M_l}(p_l)^5}{6\,(N-5)!}. \tag{A.4}$$

Hypothesis: for any $t \geq 3$,

$$\mathrm{Pe}_l(p_l)\,|_t \approx \sum_{j=3}^{t} \frac{(N-1)!\,(j-2)}{(N-j)!\,(j-1)!} P_{M_l}(p_l)^j + \frac{(N-1)! P_{M_l}(p_l)^{t+1}}{(t-1)!\,(N-t-1)!}. \tag{A.5}$$

When $t_1 = t + 1$,

$$\begin{aligned}\mathrm{Pe}_l(p_l)\,|_{t+1} &\approx \mathrm{Pe}_l(p_l)\,|_t - \frac{t+1}{N}\binom{N}{t+1} P_{M_l}(p_l)^{t+1}\left(1 - P_{M_l}(p_l)\right)^{N-t-1} \\ &\approx \mathrm{Pe}_l(p_l)\,|_t - \frac{t+1}{N}\binom{N}{t+1} P_{M_l}(p_l)^{t+1}\left(1 - (N-t-1)\, P_{M_l}(p_l)\right) \\ &= \sum_{j=3}^{t+1} \frac{(N-1)!\,(j-2)}{(N-j)!\,(j-1)!} P_{M_l}(p_l)^j + \frac{(N-1)! P_{M_l}(p_l)^{t+2}}{t!\,(N-t-2)!}.\end{aligned} \tag{A.6}$$

Therefore, the hypothesis is true for any $t \geq 3$.

*A.2 Derivation of Eq.(*3.24*)*

Consider Eq.(3.15), the first derivative of $\mathrm{Pe}_l\left(p_l\right)$ can be derived as:

$$\begin{aligned}
\mathrm{Pe}_l'\left(p_l\right) &= \frac{P_{M_l}'\left(p_l\right)}{N}\sum_{i=t+1}^{N} i^2\binom{N}{i}\left(P_{M_l}\left(p_l\right)^{i-1}\right)\left(1-P_{M_l}\left(p_l\right)\right)^{N-i} \\
&\quad -\frac{P_{M_l}'\left(p_l\right)}{N}\sum_{i=t+1}^{N} i\binom{N}{i}\left(P_{M_l}\left(p_l\right)\right)^{i}\left(N-i\right)\left(1-P_{M_l}\left(p_l\right)\right)^{N-i-1} \\
&= \frac{1}{N}\sum_{i=t+1}^{N} i\binom{N}{i}\frac{P_{M_l}'\left(p_l\right)\left(i-NP_{M_l}\left(p_l\right)\right)}{P_{M_l}\left(p_l\right)^{1-i}\left(1-P_{M_l}\left(p_l\right)\right)^{i-N+1}} \\
&= \sum_{j=t}^{N-1}\binom{N-1}{j}\frac{P_{M_l}'\left(p_l\right)\left(j+1-NP_{M_l}\left(p_l\right)\right)P_{M_l}\left(p_l\right)^{j}}{\left(1-P_{M_l}\left(p_l\right)\right)^{j-N+2}}.
\end{aligned} \tag{A.7}$$

*A.3 Derivation of (*4.22*) and (*4.23*)*

Consider the optimization problem in (4.19) with objective function $J_{\mathrm{appx}}\left(\mathbf{p}\right)$, the corresponding Lagrangian is:

$$\begin{aligned}
&L\left(\mathbf{p},\xi,\nu\right) \\
&= S\sum_{r\in\mathcal{S}} w_r a_{m_r} e^{-b_{M_r}\rho\lambda_r p_r} - \sum_{r\in\mathcal{S}}\xi_r p_r + \nu\left(\sum_{r\in\mathcal{S}} -1\right).
\end{aligned} \tag{A.8}$$

where ξ and ν are Lagrange multipliers associated with the inequality constraints and equality constraint respectively. For each $r\in\mathcal{S}$, the Karush-Kuhn-Tucker (KKT) conditions can be expressed as:

1. Primal feasible: $\mathbf{p}^*\succeq 0$; $\mathbf{1}^T\cdot\mathbf{p}^*=1$.

2. Dual feasible: $\xi^*\succeq 0$.

3. Complementary slackness: $\xi_r^*\cdot p_r^*=0$.

4. Gradient of Lagrangian vanishes:

$$\begin{aligned}&\left.\frac{\partial L\left(\mathbf{p},\xi,\nu\right)}{\partial p_r}\right|_{p_r^*,\xi_r^*,\nu^*}\\&=-Sw_r a_{m_r} b_{m_r}\rho\lambda_r e^{-b_{m_r}\rho\lambda_r p_r^*}-\xi_r^*+\nu^*=0.\end{aligned}\tag{A.9}$$

Since the optimization problem in (4.19) with objective function $J_{\text{appx}}\left(\mathbf{p}\right)$ is convex, any point satisfying the above KKT conditions is primal and dual optimal with zero duality gap [5]. The above KKT conditions imply:

$$\nu^*\geq Sw_r a_{m_r} b_{m_r}\rho\lambda_r e^{-b_{m_r}\rho\lambda_r p_r^*},\tag{A.10}$$

and

$$\left(\nu^*-Sw_r a_{m_r} b_{m_r}\rho\lambda_r e^{-b_{m_r}\rho\lambda_r p_r^*}\right)p_r^*=0.\tag{A.11}$$

If $p_r^*=0$, the approximation $J_{\text{appx}}\left(\mathbf{p}\right)$ is not valid and the r^{th} channel should be removed from $\mathcal{S}$. Thus, $p_r^*\neq 0$ in (A.11) is considered. Therefore,

$$\nu^*=Sw_r a_{m_r} b_{m_r}\rho\lambda_r e^{-b_{m_r}\rho\lambda_r p_r^*}.\tag{A.12}$$

Let $\mu=\log\left(1/\nu^*\right)$,

$$p_r^*=\frac{\mu+\log\left(Sw_r a_{m_r} b_{m_r}\rho\lambda_r\right)}{b_{m_r}\rho\lambda_r},\tag{A.13}$$

which is equivalent to (4.22). Since $\mathbf{1}^T\cdot\mathbf{p}^*=1$,

$$\mu\sum_{r\in\mathcal{S}}\frac{1}{b_{m_r}\rho\lambda_r}+\sum_{r\in\mathcal{S}}\frac{\log\left(Sw_r a_{m_r} b_{m_r}\rho\lambda_r\right)}{b_{m_r}\rho\lambda_r}=1.\tag{A.14}$$

Therefore, (4.23) is obtained by reforming (A.14). Note that the notation $\mathbf{p}^*$ used in this Appendix denotes the optimal solution of (A.8), which has different meaning to the notaion $\mathbf{p}^*$ used in (4.15) and the rest of this thesis.

VITA

Xiang Chen received the B.S. degree in electronic and communication engineering from the City University of Hong Kong, Hong Kong, in 2009 and the M.S. degree in electrical engineering from the University of Florida, Gainesville, in 2011. He is currently pursuing the Ph.D. degree in electrical engineering at University of Washington, Seattle. After his graduation, he will join Tupl Inc., a startup big data company for wireless network optimization. His research interests include multimedia networking, wireless communication, and MIMO techniques.