UNIVERSITY OF CALIFORNIA, SAN DIEGO

Cross Layer Resource Allocation For Multiple User Video Communication Systems

A dissertation submitted in partial satisfaction of the requirements for the degree
Doctor of Philosophy

in

Electrical Engineering (Communication Theory and Systems)

by

Dawei Wang

Committee in charge:

Professor Pamela C. Cosman, Co-Chair
Professor Laurence B. Milstein, Co-Chair
Professor Theodore Groves
Professor Bhaskar D. Rao
Professor Geoffrey M. Voelker

2013
The dissertation of Dawei Wang is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

________________________________________

________________________________________

________________________________________

________________________________________

Co-Chair

Co-Chair

University of California, San Diego

2013
我唯一的遗憾，是我只有一次生命献给我的祖国。
——内森·黑尔
# TABLE OF CONTENTS

Signature Page .............................................................. iii
Epigraph ................................................................. iv
Table of Contents ....................................................... v
List of Figures .............................................................. vii
List of Tables .............................................................. ix
Acknowledgements ....................................................... x
Vita ................................................................. xii
Abstract of the Dissertation ........................................ xiii

Chapter 1 Introduction .................................................. 1
  1.1 Background and Motivation ........................................ 2
  1.2 Related Work and Overview of Contributions ................. 5
      1.2.1 Related Work on Resource Allocation For Slow
            Fading Channels ........................................ 8
      1.2.2 Resource Allocation For Systems with Arbitrary
            Mobility ............................................. 10
      1.2.3 Organization of the Dissertation ....................... 12

Chapter 2 System Models .............................................. 13
  2.1 Physical Layer Model ........................................... 13
  2.2 Application Layer Model ...................................... 16
      2.2.1 Scalable Video Codec .................................. 16
      2.2.2 Nonscalable Video Codec ............................. 18
      2.2.3 Video RD Characteristics .......................... 18
  2.3 Cross Layer Resource Allocation Framework ................. 20

Chapter 3 Resource Allocation For Slow Fading Channels ........ 23
  3.1 Cross Layer Resource Allocation Problem
      Formulation ................................................ 23
  3.2 Continuous Frequency Channel Response Resource Allo-
      cation Analysis ........................................ 25
      3.2.1 Conventional Video Multiplexing ...................... 25
      3.2.2 Two-User Continuous Frequency Channel Response
            Optimal Allocation .................................. 27
  3.3 Iterative Resource Allocation Algorithm .................... 34
3.4 Baseline Algorithms ................................. 37
  3.4.1 Application Layer Optimization Algorithm .... 37
  3.4.2 Physical Layer Optimization Algorithm ......... 39
3.5 Simulation Setup ................................... 41
  3.5.1 Physical Layer Setup ............................. 41
  3.5.2 Application Layer Setup ......................... 42
3.6 Results ............................................ 43
  3.6.1 Systems with Different $SER_t$ .................. 43
  3.6.2 Systems with Different Coherence Bandwidth ... 47
  3.6.3 Complexity of Iterative Water Filling Algorithm 49
3.7 Conclusion ....................................... 50

Chapter 4 Resource Allocation For Channels With Arbitrary User Mobility 54
  4.1 Doubly Selective Channel Model ..................... 55
  4.2 Pilot Symbol Assisted Modulation and Multiple User Re-
  source Allocation .................................... 56
   4.2.1 System Operation Overview ...................... 56
   4.2.2 Pilot Symbol Assisted Modulation (PSAM) ........ 58
   4.2.3 Resource Allocation Problem Formulation ........ 61
  4.3 Resource Allocation Algorithms .................... 62
   4.3.1 Application Layer Resource Allocation Algorithm 63
   4.3.2 Physical Layer Resource Allocation Algorithm ... 64
   4.3.3 Cross Layer Resource Allocation Algorithm .... 66
  4.4 System Parameter Optimization ..................... 69
   4.4.1 System With Different $L_s$ .................... 70
   4.4.2 Systems with Different Resources ............... 73
   4.4.3 Systems With Different $\mu$ .................. 77
  4.5 System Capacity Analysis .......................... 78
  4.6 Conclusion ....................................... 81

Chapter 5 Conclusions and Future Work .................. 83

Appendix A Performance Change Calculation For Reassigning Subcarriers .... 85

Bibliography ......................................... 87
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Five-layer network architecture</td>
<td>3</td>
</tr>
<tr>
<td>1.2</td>
<td>A typical Rate Distortion curve</td>
<td>4</td>
</tr>
<tr>
<td>1.3</td>
<td>Uplink multi-carrier video communication system: video users with different</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>RD functions share a frequency band</td>
<td></td>
</tr>
<tr>
<td>1.4</td>
<td>Slow fading channel: channel stays roughly constant for the duration of GOP</td>
<td>7</td>
</tr>
<tr>
<td>1.5</td>
<td>Fast fading channel: channel varies for the duration of GOP</td>
<td>7</td>
</tr>
<tr>
<td>2.1</td>
<td>Cross-Layer optimization system transmitter diagram</td>
<td>15</td>
</tr>
<tr>
<td>2.2</td>
<td>Rate-Distortion functions for videos with different content</td>
<td>19</td>
</tr>
<tr>
<td>3.1</td>
<td>Water level change for user 1 gaining one band. The water level drops from</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td>$W_1$ to $W_1'$ after user 1 gains one additional band $\theta$.</td>
<td></td>
</tr>
<tr>
<td>3.2</td>
<td>Water level change for user 2 losing one band. The water level rises from</td>
<td>29</td>
</tr>
<tr>
<td></td>
<td>$W_2$ to $W_2'$ after user 2 loses one band $\theta$.</td>
<td></td>
</tr>
<tr>
<td>3.3</td>
<td>Uplink optimization algorithm flow chart</td>
<td>38</td>
</tr>
<tr>
<td>3.4</td>
<td>Video PSNR Performance vs. Number of Users. $SER_t = 0.2$,</td>
<td>44</td>
</tr>
<tr>
<td></td>
<td>16 Subcarriers, $\Psi=1$, Average SNR=18dB if only one subcarrier is</td>
<td></td>
</tr>
<tr>
<td></td>
<td>assigned, Users for PHY Layer Optimization are limited to be assigned at</td>
<td></td>
</tr>
<tr>
<td></td>
<td>most 1.5 times of the average remaining resources</td>
<td></td>
</tr>
<tr>
<td>3.5</td>
<td>Individual user’s performance in systems with different numbers of users.</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td>Each column indicates the same user’s RD relations in systems with four,</td>
<td></td>
</tr>
<tr>
<td></td>
<td>eight and twelve users.</td>
<td></td>
</tr>
<tr>
<td>3.6</td>
<td>Video PSNR Performance vs. Number of Users. $SER_t = 0.1$,</td>
<td>46</td>
</tr>
<tr>
<td></td>
<td>16 Subcarriers, $\Psi=1$, Average SNR=18dB if only one subcarrier is</td>
<td></td>
</tr>
<tr>
<td></td>
<td>assigned</td>
<td></td>
</tr>
<tr>
<td>3.7</td>
<td>Video PSNR Performance vs. Number of Users. $SER_t = 0.25$,</td>
<td>47</td>
</tr>
<tr>
<td></td>
<td>16 Subcarriers, $\Psi=1$, Average SNR=18dB if only one subcarrier is</td>
<td></td>
</tr>
<tr>
<td></td>
<td>assigned</td>
<td></td>
</tr>
<tr>
<td>3.8</td>
<td>Video PSNR Performance vs. $SER_t$. 4 Users, 16 Subcarriers, $\Psi=1$,</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>Average SNR=18dB if only one subcarrier is assigned.</td>
<td></td>
</tr>
<tr>
<td>3.9</td>
<td>Video PSNR Performance vs. Number of Users. $SER_t = 0.2$,</td>
<td>49</td>
</tr>
<tr>
<td></td>
<td>16 Subcarriers, $\Psi=2$, Average SNR=18dB if only one subcarrier is</td>
<td></td>
</tr>
<tr>
<td></td>
<td>assigned</td>
<td></td>
</tr>
<tr>
<td>3.10</td>
<td>Video PSNR performance degradation comparison. $SER_t = 0.2$, 16 Subcarriers,</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>$\Psi=1$ vs. $\Psi=2$, Average SNR=18dB if only one subcarrier is assigned.</td>
<td></td>
</tr>
<tr>
<td>3.11</td>
<td>Video PSNR Performance vs. Number of Users. $SER_t = 0.2$, 16 Subcarriers,</td>
<td>51</td>
</tr>
<tr>
<td></td>
<td>$\Psi=4$, Average SNR=18dB if only one subcarrier is assigned.</td>
<td></td>
</tr>
</tbody>
</table>
Figure 3.12: Video PSNR Performance Degradation Comparison. $SER_t = 0.2$, 16 Subcarriers, $\Psi=1$ vs. $\Psi=4$, Average SNR=18dB if only one subcarrier is assigned. .................................................. 52

Figure 3.13: Average MSE vs. Number of Iterations. ................................. 53

Figure 4.1: Pilot assisted modulation. One pilot symbol is added for each subcarrier to estimate the channel for every time epoch. If the number of correlated subcarriers is larger than 1, only one pilot symbol is needed for every correlated band. ................................. 58

Figure 4.2: $L_S = 100$, 16 Subcarriers, 3 Users, $SER_t = 0.15$, $\mu = 1$. The PSNR performance for a system with $f_{nd} > 10^{-2}$ is less than 15dB. .............................................................. 71

Figure 4.3: Decoded BER vs $f_{nd}$, $L_S = 100$, $SER_t = 0.15$ ........................ 72

Figure 4.4: $L_S = 25$, 16 Subcarriers, 3 Users, $SER_t = 0.15$, $\mu = 1$ ........ 73

Figure 4.5: $L_S = 5$, 16 Subcarriers, 3 Users, $SER_t = 0.15$, $\mu = 1$ ........... 74

Figure 4.6: $L_S = 100$, 16 Subcarriers, 4 Users, $SER_t = 0.15$, $\mu = 1$ ...... 75

Figure 4.7: $L_S = 100$, 24 Subcarriers, 3 Users, $SER_t = 0.15$, $\mu = 1$ ....... 75

Figure 4.8: $L_S = 100$, 24 Subcarriers, 4 Users, $SER_t = 0.15$, $\mu = 1$ ....... 76

Figure 4.9: $f_{nd} = 10^{-3}$, $L_S = 100$, 16 Subcarriers, $SER_t = 0.15$ ........ 78

Figure 4.10: $f_{nd} = 10^{-3}$, $L_S = 25$, 16 Subcarriers, $SER_t = 0.15$ ........ 79

Figure 4.11: $L_S = 100$, $f_{nd} = 10^{-6}$, 16 Subcarriers, $SER_t = 0.15$, $\mu = 1$ 80

Figure 4.12: $L_S = 25$, $f_{nd} = 10^{-3}$, 16 Subcarriers, $SER_t = 0.15$, $\mu = 1$ 81
## LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1.1</td>
<td>Comparison of wireless traffic</td>
<td>3</td>
</tr>
<tr>
<td>Table 1.2</td>
<td>Resource allocation in different layers</td>
<td>10</td>
</tr>
<tr>
<td>Table 4.1</td>
<td>Evolution of the performance: 16 Subcarriers/3 Users</td>
<td>76</td>
</tr>
<tr>
<td>Table 4.2</td>
<td>Evolution of the performance: 24 Subcarriers/3 Users</td>
<td>77</td>
</tr>
</tbody>
</table>
ACKNOWLEDGEMENTS

There are a lot of people I wish to thank for making my experience as a graduate student at UCSD the most memorable period in my life so far. First of all, I would like to thank my advisors Prof. Cosman and Prof. Milstein for their support and guidance over the past years. Prof. Cosman always sees the big picture and shows great patience and faith in my ability to allow me to work independently. Her broad knowledge on different research fields help me to organize the fundamentals of this dissertation and her brilliant technical writing skills help me to express my ideas accurately and clearly. Prof. Milstein is not only a sincere mentor but also a role model. He is always willing to share his encyclopedia-like knowledge on communications and his attitude towards research and life is one of the most valuable things I learned in UCSD.

I am grateful to Prof. Theodore Groves, Prof. Bhaskar Rao and Prof. Geoffrey Voelker for serving as members of my dissertation committee. Their invaluable suggestions and comments help to strengthen this dissertation.

I am also deeply indebted to current and past members of my lab, Yueh-Lun Chang, Hobin Kim, Ting-Lan Lin, Tu Nguyen, Mehdi Saleh, Qing Song, Madushanka Soysa, Vanessa Testoni, Mayank Tiwari, Arash Vosoughi and Yuxia Wang, to name just a few. Ting-Lan and Mayank’s guidance on video coding really helped me to start my simulations early in my research. I appreciate Laura Toni in particular for her help and inspirations in a crucial period of my work. Laura is a constant source of knowledge, help and most importantly, confidence.

I would also like to thank my friends, including Siyuan Gu, Feng Lu, Qihang Peng, Jing Wang and Junni Zou, for making the years at UCSD fun and enjoyable.

Finally, I am most grateful to my parents, without whom none of this would be possible. I can’t imagine traveling across the Pacific Ocean to study in California for such a long time without their love and support.

2060-2073. The dissertation author was the primary investigator and author of this paper.

Chapter 4, in part is currently being prepared for submission for publication by Dawei Wang, Laura Toni, Pamela Cosman and Laurence Milstein "Resource Allocation and Performance Analysis for Multiuser Video Transmission over Doubly Selective Channels", *IEEE Transaction on Wireless Communications*. The dissertation author is the primary investigator and author of this paper.
VITA

1985 Born in the City of Ningbo, Zhejiang, China

2008 B. Eng. in Electronic Engineering, Hong Kong University of Science and Technology, Kowloon, Hong Kong SAR, China

2011 M.S., in Electrical Engineering (Communication Theory and Systems), University of California, San Diego

2013 Ph. D. in Electrical Engineering (Communication Theory and Systems), University of California, San Diego

PUBLICATIONS


ABSTRACT OF THE DISSERTATION

Cross Layer Resource Allocation For Multiple User Video Communication Systems

by

Dawei Wang

Doctor of Philosophy in Electrical Engineering (Communication Theory and Systems)

University of California, San Diego, 2013

Professor Pamela C. Cosman, Co-Chair
Professor Laurence B. Milstein, Co-Chair

In this dissertation, we investigate problems of resource allocation for uplink multiuser wireless communication channels. The users want to upload compressed video streams with different rate-distortion (RD) characteristics to a base station via a multi-carrier channel. The video is organized in groups of pictures (GOP), and the base station makes resource allocation decisions based jointly on the channel state information (CSI) and the RD information to minimize the average video distortion across all users.

We first study the case where the channel is slowly varying. With the as-
sumption of the channel staying constant over the duration of a GOP, we derive an optimal condition for minimizing the sum of distortions in a setting of continuous frequency channel response. Our result shows that for a specific band to be assigned to a particular user, the product of the slope of the RD curve of that user and a marginal rate increment corresponding to that band should be maximized. We then design an algorithm which iteratively calculates the status of the RD information and the CSI in a block fading setting. Compared to two baseline resource allocation algorithms that use only a single layer of information, the simulation results show that the cross layer scheme can consistently improve the performance of the system.

We then focus on the case of a channel with arbitrary mobility. To better utilize the time varying channel, we design a pilot symbol assisted modulation (PSAM) system which periodically adapts the modulation format according to the instantaneous CSI. We show that the two baseline algorithms with single layer information behave quite differently as the Doppler spread changes, but the cross layer algorithm performs robustly across all different mobile speeds. We further characterize the capacity gain of using the cross layer scheme.
Chapter 1

Introduction

Over the past two decades or so, the wireless communication industry has been experiencing a spectacular technological revolution. The development of physical (PHY) layer transmission schemes, such as Multiple Input and Multiple Output (MIMO) antennas, as well as multiple access schemes, e.g., Orthogonal Frequency Division Multiple Access (OFDMA), have profoundly changed everyday life. The combination of ubiquitous Wireless Local Area Network (WLAN) with cellular phone coverage creates an environment where users desire seamless, high quality connectivity at all times, with all different mobile speeds, and from virtually all locations.

Meanwhile, over the past ten years, new video compression techniques have dramatically improved the compression efficiency of video codecs. The updated H.264 Advanced Video Coding (H.264/AVC) video coding standard [1] has been shown to save 50% of the bits compared to prior MPEG standards. Most recently, the new High Efficiency Video Coding (HEVC) [2] is poised to further improve over the coding efficiency of the H.264 codec.

With the maturity of the technologies at both the physical layer and the application layer, we recently witnessed an explosion of demand of video traffic over wireless communication networks. Cisco predicts that, by 2015, various forms of video data will represent over 90% of global data communication traffic, and almost 66% of the world’s mobile traffic [3]. While system optimization at each
single layer continues to develop, more attention has been paid to optimizing system performance by jointly considering the application layer and physical layer information, i.e., cross layer optimization.

1.1 Background and Motivation

Traditionally, communication systems are designed in layers, with each layer corresponding to a specific functionality of the communication task. A conventional five-layer system architecture is shown in Fig. 1.1 [4]. Information data generated from the source side will normally be packetized. At each layer, a packet header, which contains the key layer information, will be added to each packet [5]. For example, at the network layer, an Internet Protocol (IP) address will be added before packets enter the network. With overhead added at each layer, the layered structure significantly decreases the end-to-end efficiency. More importantly, in order to simplify the system design, the information of a higher layer is normally abstracted or veiled to a lower layer, and the characteristics of higher layers (particularly the application layer) are not considered in lower layer design, so the overall system performance is not optimized.

One of the best examples of the inefficiency of the layered structure is video communication over wireless and wireline systems. A table of the source characteristics for three of the main communication sources is listed in Table 1.1 [6]. Compared with an audio source, the video source normally needs three orders of magnitude higher data rate, and the nature of the compressed video requires a much smaller bit error rate. Compared to data sources, the video source is delay sensitive. Robust and continuous transmission at different channel conditions are highly desired for a satisfactory user experience.

Besides the difference with other sources, the compressed video data itself usually displays different characteristics. Within one video sequence, encoded bits of information for motion vectors and macroblock (MB) ID’s are more important than bits such as the least significant bit of a pixel value. However, bits are usually equally protected in the transmission, because in a layered system, such
importance information is not available for the lower layers. For different video sequences, the video characteristics are highly content-dependent. For the same resolution video, the number of bits required to achieve the same video quality is
closely related to the characteristics of the video frame and the encoding structure of the sequence. These content dependent characteristics are normally described by a Rate-Distortion (RD) function. A typical RD function is shown in Fig. 1.2, where the distortion is measured in Mean Square Error (MSE). In the literature, video rate control algorithms as well as RD analysis were studied extensively [7–11]. These papers show that the RD functions are normally strictly convex and decreasing, and can be dramatically different depending on the complexity (high or low motion) of a video stream. The details about the model for an RD function will be discussed in Chapter 2.

![A typical Rate Distortion curve.](image)

**Figure 1.2**: A typical Rate Distortion curve.

Recently, the emerging demand for video transmission and the limitation of communication bandwidth have become the bottleneck to further development of industry. The paradigm of cross layer design breaks the boundary line of layers and jointly optimizes the end-to-end performance of the system. Some attempts at cross layer design for video communications have been used in commercial products. Video oriented protocols such as HTTP adaptive streaming (HAS) and dynamic adaptive streaming over HTTP (DASH) have been introduced for delivering enhanced quality of experience (QoE) video [12] [13]. However, in most communication systems, especially wireless communication systems, video traffic
is still mixed with other traffic, and the characteristics of the video are not used in lower layer design.

For cellular and wireless local area network systems, research on various PHY/MAC/APP cross layer techniques has improved video quality and increased network capacity. Cross layer techniques include: single user point-to-point PHY/APP optimization and multiuser multiple access optimization.

Point-to-point PHY/APP cross layer optimization seeks to exploit the uniqueness of the video data, and schemes such as multiple description coding [14] [15], unequal error protection (UEP) [16] [17] and joint source channel coding [18] [19] have been developed. The results of these works show that the quality of received video can be significantly enhanced by protecting the video data with different levels according to the relative importance of the bits.

In this dissertation, we focus on cross layer optimization in a multiple access environment. We investigate the resource allocation strategy by jointly considering the application layer information and the physical layer information. With the goal of optimizing the overall system video performance, the physical layer communication resources (bandwidth and power) are allocated to the video users based on the demand of the video user and the characteristics of the channel.

1.2 Related Work and Overview of Contributions

In this dissertation, we consider a PHY/APP cross layer resource allocation technique for a multiple video user uplink system (Fig. 1.3). The users share a common wireless band and the RD functions are potentially different from user to user. The base station collects both the RD functions and the Channel State Information (CSI), and allocates bandwidth according to the demand of video users to optimize the overall performance of the system. In most of the dissertation, the objective of the optimization problem will be minimizing the sum of distortions among the video users.

At the application layer, the video is organized and compressed in units of GOPs, and the measure of the objective (distortion) is calculated GOP by GOP
as a function of the encoding rate. At the physical layer, multi-carrier system is a flexible and low-complexity way of managing communication resources [20] [21]. By dividing a large piece of bandwidth (presumably a frequency selective channel) into narrowband subcarriers, the multi-carrier system provides the resource allocator the freedom to exploit multi-user diversity (MUD) in different frequency bands. In this dissertation, we treat the subcarrier bandwidth and transmission power as the resource pool for cross layer resource allocation, and design allocation schemes jointly based on the video characteristics (RD function) and the CSI of the subcarriers. Note that the transmission power is normally constant by design, so the dynamics of the resource pool in a cross layer resource allocation problem is typically dependent on the dynamics of the CSI of subcarriers. We will investigate
two different cases under our cross layer resource allocation framework, the case where CSI is stable for the duration of a GOP (Fig. 1.4) and the case where CSI is changing rapidly (Fig. 1.5).

Figure 1.4: Slow fading channel: channel stays roughly constant for the duration of GOP

Figure 1.5: Fast fading channel: channel varies for the duration of GOP
1.2.1 Related Work on Resource Allocation For Slow Fading Channels

When the dynamics of the resource pool (CSI of the channel) are varying slowly within the GOP duration, the amount of throughput allocated to users can be estimated accurately when the resource allocation decision is made. We note that the resource allocation design for a single layer design (application layer only or physical layer only resource allocation) has been well studied in the literature, and most work has focused on the case of slow dynamics.

At the physical layer, the problem of assigning resources in a multi-carrier system was studied in [22], where the authors formulated and solved a total transmission power minimization problem for different user quality-of-service (QoS) requirements. Research in [23–27] tried to solve the throughput maximization problem, given power and spectrum constraints in different communication settings. Because of the complexity of the optimization problems, most of the work above proposed numerical algorithms instead of finding analytical solutions. Power allocation for an imperfect CSI case was explored in [28, 29]. To reduce the complexity of resource allocation algorithms, chunk-based resource allocation, which makes allocation decisions on subcarriers in groups, was studied by [30, 31]. Results show that when the chunk bandwidth is smaller than the coherence bandwidth, the chunk-based resource allocation can significantly reduce the computational complexity while maintaining similar throughput performance compared to subcarrier-based resource allocation algorithms. Utility driven resource allocation was investigated by [32,33] and most recently by [34,35] in an information theoretic setting. In these papers, instead of maximizing the sum of the throughputs, the objective of the optimization is the overall utility, which is a function of throughput. For all the research listed here, the objective metrics (power consumption or throughput) are calculated for the duration in which the CSI stays unchanged.

At the application layer, since the rate-distortion tradeoff will be highly dependent on the content of the video [36,37], the diversity of different video RD curves provides us an opportunity to optimize the overall video quality when mul-
tiple video streams share the same resource pool, i.e., video multiplexing. For most of the literature on video multiplexing, the resource pool is either a fixed number of bits or a fixed bit rate for the duration of resource allocation, and the authors assume an error-free scenario. In [38], the authors considered a multiple camera surveillance system, and exploited the difference between high complexity and low complexity videos. In [39], the economic concept of competitive equilibrium is used to allocate bit rate. The authors show that, with a fixed sum of bit rate for the duration of GOPs, the video quality of each individual user improves by trading bit rate between users across time. Again, in the work above, the amount of the resource (bits or bit rate) is assumed to be constant over the duration that the distortion is measured. When multiplexing videos in a wireless mobile communication case, bit rates will be determined by the available bandwidth, transmission power and CSI. In this sense, multiplexing video streams in a wireless environment with a resource pool of power and bandwidth will be more challenging than conventional video multiplexing. In a cellular wireless multi-carrier video transmission system, the CSI as well as the complexity of video streams can be collected by the base station. Both multiuser channel diversity and video complexity diversity can be used simultaneously to optimize the power and subcarrier assignments. In [40], the authors propose a joint uplink and downlink cross layer resource allocation framework with the resource being the channel access time duration. In [41] and [42], the authors study a subcarrier and power assignment problem in a downlink setting, where the subcarrier assignment and power allocation are treated as two independent steps. To better optimize the system, we propose an iterative algorithm which allows the application layer and physical layer to interact. In Chapter 3, we are interested in a cooperative setting for a slow fading scenario, where the goal is to minimize the sum of distortions among users. The distortion is measured in the unit of GOPs, and the channel is assumed to be constant within each duration (see Fig. 1.4). Video streams with high complexity should be given more subcarriers with good channel gains, while streams with low complexity will get a relatively small number of subcarriers.

A table of resource allocation schemes at different layers can be found in
Table 1.2. For resource allocation at the application layer (video multiplexing), the allocator has a fixed budget of bits or bit rate and the task is to allocate the bits (or bit rate) to the users such that the sum of distortions is minimized. In a scenario where only physical layer information is involved, the goal of the allocator is to maximize the sum of throughputs via assigning power and bandwidth to different users. In this dissertation, we will consider the case where the resource pool is power and bandwidth, and the objective is to minimize the sum of distortion. Since both the application layer information and the physical layer information are user dependent, the optimization problem for a cross layer problem will be much more complicated than that for a single layer optimization.

<table>
<thead>
<tr>
<th>Layer</th>
<th>Resource Pool</th>
<th>Objective</th>
</tr>
</thead>
<tbody>
<tr>
<td>Application Layer</td>
<td>Bits or Bit Rate</td>
<td>Minimize Sum of Distortion</td>
</tr>
<tr>
<td>Physical Layer</td>
<td>Power and Bandwidth</td>
<td>Maximize Sum of Throughput</td>
</tr>
<tr>
<td>Cross Layer</td>
<td>Power and Bandwidth</td>
<td>Minimize Sum of Distortion</td>
</tr>
</tbody>
</table>

1.2.2 Resource Allocation For Systems with Arbitrary Mobility

When a mobile user in the system experiences high mobility, the channel realization will vary rapidly with respect to time. In Fig. 1.5, we show a system where the coherence time of the channel is smaller than the duration of one GOP. To combat the uncertainty of the time-varying channel, video communication with Automatic Retransmission reQuest (ARQ) was proposed in [43] and [44]. Although ARQ is easy to implement, the time delay and uncertainty for exchanging the ARQ signals might not be suitable for delay sensitive video applications. More importantly, for a system with different Doppler spreads, the packet loss rate (PLR) varies dramatically with respect to channel estimation accuracy, which is determined jointly by the pilot spacing, the pilot-to-data power ratio and the number of...
pilots used for interpolation [45]. We found that for most papers studying ARQ-based video communication, the PLR model is normally oversimplified. In [43], PLR is treated as a constant for all Doppler spreads. In [44], the authors study the performance of adaptive modulation with ARQ in a data communication system, and perfect channel estimation is assumed for choosing the modulation format and demodulation at the receiver. Channel estimation accuracy could be improved by reducing the interval between the pilots. However, the throughput loss due to pilot insertion might significantly reduce the number of video source bits delivered to the channel. Under the perfect CSI assumption, the critical tradeoff between the channel estimation accuracy and source encoding rate is missed in [44].

Forward Error Correction (FEC)-based video communication with no retransmission is often used for delay-sensitive video data. To achieve higher average image quality in systems with high mobility, [46] utilizes the coding diversity across both time and frequency, and analyzes the performance of progressive image transmission in the presence of inter-carrier interference and channel estimation error in a multi-carrier setting. In [47], the authors study a joint link and source adaptation system, where the modulation and coding scheme at the PHY layer is chosen according to instantaneous CSI and different importance levels of the packet, while the source rate at the APP layer is chosen based on the visibility of the packet and playback buffer status. Without any assumption of knowledge of the channel in the future, the adaptation scheme in [47] has the potential to be applicable for systems with arbitrary mobility. In [48], a rate compatible punctured convolutional (RCPC) code was combined with a Reed-Solomon code to protect image data against fast channel variation, and simulation results show that the product code scheme performs robustly for both a Gilbert-Elliot two-state channel model and the Jakes’ model [49].

We find that the problem of resource allocation design and performance analysis for systems with arbitrary mobility is relatively under-investigated. With the channel quality varying rapidly, the transmission strategy should be adaptive across time. On the other hand, because of the complexity of the resource allocation and the overhead of the information exchange, the resource allocation might
have to be done over a longer time scale than the coherence time.

1.2.3 Organization of the Dissertation

We start this dissertation by introducing the physical layer and application layer settings for the resource allocation problem in Chapter 2. In Chapter 3, we will focus on the resource allocation problem for channels with low mobility. As discussed in the previous section, when the channel stays constant for the duration of the resource allocation unit (GOP duration), the effect of pilot overhead and channel estimation error will be negligible. We then extend the analysis to resource allocation for channels with arbitrary mobility in Chapter 4. We investigate the tradeoff between channel estimation accuracy and pilot throughput loss extensively in this chapter. We draw conclusions and propose future work in Chapter 5.
Chapter 2

System Models

In this chapter, we introduce the system models at both the physical layer and application layer. The general models described in this chapter will be used in both Chapter 3 and Chapter 4, however specific details of models will differ in different parts of the dissertation, and will be discussed more carefully when needed.

We consider a cellular multi-carrier video communication system with a set of video users \( k = \{1, 2, 3 \ldots K\} \). The system occupies a total frequency band of \( W \) (Hz) equally divided into \( M_c \) orthogonal subcarriers \( m = \{1, 2, 3 \ldots M_c\} \). Let \( f_{nd}^k \) be the normalized Doppler spread of user \( k \). For most of the systems in Chapters 3 and 4, we will drop the index \( k \) and assume that users experience the same Doppler spread. We focus on an uplink system, and the task of the resource allocation is to allocate subcarriers to the users based on both RD functions and CSI for different channel settings.

2.1 Physical Layer Model

The system operates in a slotted manner, and the length of one time slot is \( T_s \) (sec) for both downlink and uplink. The slot time \( T_s \) is equal to the video display time of one GOP, and the transmission duration for each GOP spans one time slot. For all \( K \) subcarriers, we assume that the channel gain within each
subcarrier is flat. In the resource allocation, each subcarrier can only be used by one user, but it is possible for one user to get more than one subcarrier. Let 

$$H_s^k[l] = [H_{s,1}^k[l], H_{s,2}^k[l], \ldots H_{s,m_s}^k[l], \ldots H_{s,M_c}^k[l]]$$

be the complex channel gain of user \(k\) for the set of subcarriers at the \(l\)-th symbol of time slot \(s\). The subcarrier assignment as well as the power allocation decision will be made on a slot-to-slot basis.

A popular implementation of the multi-carrier is the orthogonal frequency division multiplexing (OFDM) system. A block diagram of the transmitter of an OFDM system is shown in Fig. 2.1. Let \(T\) be the data duration and \(T_{cp}\) be the length of the cyclic prefix. We define \(T_0 = T + T_{cp}\) to be the duration of an OFDM symbol. The baseband transmitted signal for user \(k\) can be written as

$$x_k(t) = \sum_l \sum_{m=1}^{M_c} \sqrt{P_{k,m}} X_{k,m}[l] \exp\left(\frac{j2\pi mt}{T}\right) \Pi(t - lT_0) \quad (2.1)$$

where \(P_{k,m}\) and \(X_{k,m}[l]\) are the transmission power and coded symbol with unit variance respectively of user \(k\) on subcarrier \(m\). We drop the GOP index \(s\) for simplicity here, and \(P_{k,m}\) is assumed to be fixed for the duration of a time slot. \(P_{k,m} = 0\) if subcarrier \(m\) is not allocated to user \(k\). Also, \(\Pi(t) = 1, \forall t \in [0, T_0)\), and \(\Pi(t) = 0\) otherwise.

Since we assume flat fading for each subcarrier, \(H_{k,m}[l]\) is a constant within a subcarrier, and the lowpass equivalent received signal of user \(k\) on subcarrier \(m\) is given by

$$y_{k,m}(t) = \sqrt{P_{k,m}} H_{k,m}[l] X_{k,m}[l] \exp\left(\frac{j2\pi mt}{T}\right) + n_{k,m}(t) \quad (2.2)$$

where \(n_{k,m}(t)\) is Additive White Gaussian Noise (AWGN) with two-sided power spectral density \(N_0\).

To detect the signal on subcarrier \(m\), a correlation operation is performed:

$$Y_{k,m} \triangleq \frac{1}{T} \int_0^T y_{k,m}(t) \exp(-j2\pi mt/T)dt \quad (2.3)$$

The noise power is given by \(P_N = E[|N_{k,m}|^2] = 2N_0/T\), and the power for the desired signal is \(P_{k,m}|H_{k,m}[l]|^2\). If the modulation format is adaptive QAM, from [50]
Figure 2.1: Cross-Layer optimization system transmitter diagram
and [51], the symbol error rate (SER) for an AWGN channel can be approximated as

\[
SER \approx 4Q \left( \sqrt{\frac{3}{M - 1} \frac{P_{k,m} |H_{k,m}[l]|^2}{P_N}} \right)
\]  

(2.4)

Here, \( \hat{M} \) is the alphabet size of a QAM waveform and for a given fixed \( SER_t \), the information rate (number of bits each symbol can carry) \( R_{k,m}(P_{k,m}, H_{k,m}[l]) \) (in bits/symbol) can be written as a function of transmission power and channel response gain:

\[
R_{k,m}(P_{k,m}, H_{k,m}[l]) = \min\{\lceil \log_2 \left( 1 + \eta P_{k,m} |H_{k,m}[l]|^2 \right) \rceil, R_{max} \}
\]  

(2.5)

where \( \eta = \frac{3}{P_N} [Q^{-1}(SER_t/4)]^{-2} \) and \( R_{max} \) is the largest alphabet size the system allows. The bit rate (in bits/sec) then can be written as \( R_{k,m}(P_{k,m}, H_{k,m}[l])/T_0 \). In the following sections, we will replace \( H_{k,m}[l] \) by the estimate \( \bar{H}_{k,m}[l] \) in (2.5) and use \( R_{k,m}(P_{k,m}, \bar{H}_{k,m}[l]) \) to determine the modulation alphabet size. The effect of the channel estimation accuracy and the choice of \( SER_t \) will be investigated in the simulation sections of Chapters 3 and 4.

2.2 Application Layer Model

We assume that the video data for all \( K \) users are compressed one GOP at a time, and the number of pictures \( (R_p) \) for one GOP is the same across the users. The frame rate for the videos is \( R_f \) (frames/second) and each GOP occupies a time slot \( T_s = R_p/R_f \) to deliver the video. As illustrated in Fig. 2.1, the number of bits transmitted in each time slot is decided by the cross layer resource allocation decision.

2.2.1 Scalable Video Codec

A scalable video codec is designed such that the decoder only needs a portion of the encoded bitstream (a substream) to display the entire video. The
decoded fidelity of the video will depend on the length and importance of the video substream, as well as the rate distortion characteristics of the video content. The flexibility of the scalable video codec is considered a powerful feature to allow adaptation to the time varying nature of the wireless channel and the throughput variations in multiple-hop communication systems.

For each bitstream, the most important video information (e.g., motion vector, frame index, macroblock ID’s) is contained in a substream called the Base Layer (BL). One or more enhancement layers (EL) are added such that the Mean Square Error (MSE) will decrease when more enhancement bits are received by the decoder. Previously, a fine granular-scalability (FGS) codec was proposed [52] [53] for accurate source-to-channel adaptation. FGS allows every successfully delivered video bit to improve the video quality, but the granularity of the scalability will significantly sacrifice the video compression efficiency.

Recently, the scalable extension of the new H.264/AVC standard (known as H.264/SVC) [54] with medium granular-scalability (MGS) has emerged as a balanced solution for the tradeoff between the compression efficiency and the scalable granularity. An H.264/SVC MGS codec features temporal, spatial and quality scalability, and allows the flexibility of dropping a combination of substreams according to the communication channel. In this dissertation, we are interested in the quality scalable function, which packetizes the encoding bitstream according to the zonal location of the DCT coefficients and ranks the packets based on their importance in the GOP. The encoder would then assign the highest priority for transmission to the packets which can most effectively reduce the compression distortion. If an error occurs in the transmission, the entire packet and all the other packets with lower priority will be dropped, but all the previous packets with higher priority which were successfully received by the decoder will be used for decoding the video.

When the channel experiences time selective fading with arbitrary Doppler spread, the number of bits that the channel can support (with some QoS guarantee) cannot be predicted with just the knowledge of the channel at the beginning of a GOP. In this sense, the scalable codec is suitable for a time-varying channel, and
we will use the H.264/SVC video codec for Chapter 4.

2.2.2 Nonscalable Video Codec

Unlike a scalable codec, a nonscalable codec requires the decoder to receive the entire bitstream to display the full resolution video content. In other words, every macroblock (or video slice) is compressed to a unique bitstream and can only be reconstructed at the decoder side by successfully receiving the corresponding bitstream. Channel error and packet loss will cause some video content (frame or macroblock) loss and is normally handled by error concealment methods [55] [56].

In a multiuser video system (Fig. 2.1), if a nonscalable codec is adopted at the application layer, the encoding rate is determined by the channel condition (CSI, bandwidth, etc) and the cross layer resource allocation algorithm decision. While being efficient in compression, a nonscalable codec normally requires accurate prediction of the channel throughput at the beginning of transmission or a sophisticated buffering algorithm to guarantee that most of the bits generated by the encoder will be delivered to the receiver. The development of rate control and rate distortion algorithms ([11] [57] [58]) show that the target video encoding rate can be precisely achieved while the distortion of the video can be optimized. In practical cases, raw video is normally compressed at different rates, and the transmitter can choose to deliver one bit stream according to the channel conditions using the stream switching approach. In Chapter 3, we assume that the encoding time of the nonscalable codec is negligible and the video can be compressed at any rate based on the resource allocation decision.

2.2.3 Video RD Characteristics

For both scalable and nonscalable video codecs, the rate distortion function characterizes the tradeoff between the compression fidelity and the number of bits used to describe the source. Since the video is compressed in units of GOPs, this RD function is also measured on a GOP-to-GOP basis. Let $D_k^s(B)$ be the rate distortion function of user $k$ in time slot $s$. For a nonscalable codec, $B$ is the
number of bits the encoder generated, and for scalable codec, $B$ is the number of bits in the substream (the length of the truncated bit stream). For each GOP, the MSE distortion can be approximated as [36]:

$$D^s_k(B) = a_k + \frac{w_k}{B + v_k} \quad (2.6)$$

where $a_k$, $v_k$ and $w_k$ are constants which depend on the video content. For a video with low complexity (e.g., News Anchor sequence in Fig. 2.2), the time and spatial redundancy can be easily compressed, since the picture is relatively uniform in area and the motion of the video is slow, and one would expect a relatively flat RD function. For a video with high complexity (e.g., high motion video such as the Mobile sequence in Fig. 2.2), the RD function is normally steep and $w_k$ is relatively large. The difference of the RD tradeoff between different users constitutes application layer diversity, which will be jointly exploited with the physical layer diversity in the cross layer design.
2.3 Cross Layer Resource Allocation Framework

In an uplink multi-carrier system (Fig. 1.3), the mobile unit submit the application layer RD values \((a_k, v_k, w_k)\) of the current GOP in their buffers. Pilot symbols are sent from the users to the base station. The resource allocator at the base station collects the estimated channel gain for each user at each subcarrier for the first symbol of the GOP \(H_k[1] = [\tilde{H}_{k,1}[1], \tilde{H}_{k,2}[1], \ldots, \tilde{H}_{k,Mc}[1]]\), for resource allocation purposes. The estimation accuracy depends on the Doppler spread of the system, \(f_{nd}\), the estimation algorithm, and the power of the pilot symbol. The resource allocation decision will then be fed back to the users. The source encoding strategy and the power allocation scheme at the transmitter of the users (Fig. 2.1) will be based on the resource allocation decision.

Based only on the CSI at the beginning of the GOP, the base station estimates the throughput of each subcarrier using (2.5), assuming that the adaptive QAM format will last for the duration of \(T_s\). The number of bits transmitted over subcarrier \(m\) of user \(k\) can then be written as \(R_{k,m}(P_{k,m}, \tilde{H}_{k,m}[1]) \cdot T_s/T_0\), where \(T_s/T_0\) is the number of QAM symbols for a GOP. We denote by \(\lambda\) (\(0 < \lambda \leq 1\)) the fraction of data symbols and \([(1 - \lambda) \cdot T_s/T_0]\) symbols will be used as pilots for channel estimation. To protect the data, a channel code of fixed rate \(u\) is added. If the channel stays constant as \(\tilde{H}_{k,m}[1]\), the information bits that the physical layer can support for user \(k\) across all \(Mc\) subcarriers is given by

\[
B_k = \left\lfloor \sum_{m=1}^{Mc} u \cdot \lambda \cdot R_{k,m}(P_{k,m}, \tilde{H}_{k,m}[1]) \cdot T_s/T_0 \right\rfloor \tag{2.7}
\]

with \(P_{k,m} = 0\) if subcarrier \(m\) is not assigned to user \(k\).

For the resource allocation algorithm design, we ignore the effect of channel errors and assume that the modulation format is constant for the GOP duration. In Chapters 3 and 4, we use (2.7) as the channel throughput for our mathematical analysis and algorithm design. For a system with large Doppler spread, the modulation format could change over the duration of a GOP, and the effect of channel errors and channel estimation errors will have a significant impact on the decoder performance. All these effects will be evaluated by simulations in the following...
chapters.

If we plug (2.7) into (2.6), then the MSE distortion for user $k$ can be written as

$$D_k = a_k + \frac{b_k}{\sum_{m=1}^{M_c} \lambda R_{k,m}(P_{k,m}, \tilde{H}_{k,m}[1]) + c_k}$$

(2.8)

Here, we have divided both the numerator and denominator by $u \cdot T_s/T_0$ for simplicity. So

$$b_k = \frac{w_k}{(u \cdot T_s/T_0)}$$

$$c_k = \frac{v_k}{(v \cdot T_s/T_0)}$$

(2.9)

With the knowledge of the physical layer CSI ($\tilde{H}_k[1]$), and the application RD ($a_k$, $b_k$ and $c_k$), the base station needs to assign $M_c$ subcarriers to $K$ users at the beginning of each GOP (time slot), and users can access the subcarrier for the duration of the time slot (GOP duration). The allocation decision will be updated at the beginning of the next GOP as both CSI and RD are updated. Mathematically, our resource allocation goal is to minimize the sum of distortions among $K$ users at each time slot $s$. The optimization objective is

$$\min_P \sum_{k=1}^{K} \frac{b_k}{\sum_{m=1}^{M_c} \lambda R_{k,m}(P_{k,m}, \tilde{H}_{k,m}[1]) + c_k}$$

(2.10)

where

$$P = \begin{bmatrix}
P_{1,1} & P_{1,2} & \ldots & P_{1,M_c} \\
P_{2,1} & P_{2,2} & \ldots & P_{2,M_c} \\
\vdots & \vdots & \ddots & \vdots \\
P_{K,1} & P_{K,2} & \ldots & P_{K,M_c}
\end{bmatrix}$$

(2.11)

is the power allocation matrix whose entry in the $k$-th row and $m$-th column, $P_{k,m}$, is the power allocation of the $m$-th subcarrier for user $k$. The base station then sends the resource allocation decision back to the users. We drop the $a_k$ term, as it is constant with respect to $P$.

We assume that any subcarrier can be used by one user exclusively, so one of the feasibility constraints for the optimization problem is: for $m \in \{1, 2, 3...M_c\}$, if $\exists k'$ such that $P_{k',m} \neq 0$, then $P_{k,m} = 0$, $\forall k \neq k'$. 
Note that the normalized Doppler spread $f_{nd}$ is not a parameter in the resource allocation formulation. In Chapter 3, we will first investigate a case where the channel is stable (small $f_{nd}$) for the duration of a GOP. We will generalize to arbitrary $f_{nd}$ in Chapter 4. Additional feasibility constraints on power will be discussed later in different chapters.
Chapter 3

Resource Allocation For Slow Fading Channels

In this chapter, we study the case of resource allocation for a slow fading channel. When the channel stays relatively constant within the duration of the GOP, the channel throughput for the duration of the GOP can be calculated accurately and the source encoder can adjust to the channel response. In the simulation, we use a nonscalable encoder and the encoding rate is controlled by the aggregate rates across all subcarriers.

3.1 Cross Layer Resource Allocation Problem Formulation

In the following sections, we assume that $f_{nd} = 0$ for every slot, or in other words, the channel coefficient for the subcarrier $m$ of user $k$ satisfies $H_{k,m}^s[1] = H_{k,m}^s[2] = ...H_{k,m}^s[T_s/T_0]$, and we will drop the GOP index $s$ and time index $l$ in this chapter. In an uplink multi-carrier system (Fig. 1.3), the base station collects the RD values ($a_k$, $b_k$, and $c_k$) and estimated CSI ($\tilde{H}_{k,m}, k = \{1, 2...K\}, m = \{1, 2...M_c\}$) as the input of the optimization problem defined in (2.10). In this chapter, as the channel is varying slowly, we assume a perfect channel estimation, or $\tilde{H}_{k,m} = H_{k,m}$ for $k = \{1, 2...K\}, m = \{1, 2...M_c\}$. In addition, we have $\lambda = 1$,
meaning that the number of pilots used for estimation is negligible. Our resource allocation problem then becomes:

$$\min_{P} \sum_{k=1}^{K} b_k \frac{\sum_{m=1}^{M_c} R_{k,m}(P_{k,m}, H_{k,m}) + c_k}{\sum_{m=1}^{M_c} P_{k,m}} \quad (3.1)$$

where $P$ is the power allocation matrix whose entry in the $k$-th row and $m$-th column, $P_{k,m}$, is the power allocation of the $m$-th subcarrier for user $k$.

We assume that each user has a total power constraint of $P$ over different subcarriers, and any subcarrier can be used by one user exclusively. The feasible solutions for this problem satisfy the following two constraints:

(C1) For $m \in \{1, 2, 3...M_c\}$, if $\exists k'$ such that $P_{k',m} \neq 0$, then $P_{k,m} = 0, \forall k \neq k'$

(C2) $\sum_{m=1}^{M_c} P_{k,m} \leq P \forall k \in \{1, 2, 3...K\}$

Since the channel coefficient is constant for the duration of GOP, the optimal power allocation (in the throughput maximization sense) calculated at the resource allocation will still be optimal for the duration of the GOP. For the optimization problem defined in (3.1), since constraint C1 is not a convex set, and this optimization problem is NP-hard, we propose an algorithm for a sub-optimal solution with two steps:

Step 1: The base station assigns subcarriers to different users according to channel conditions and rate-distortion curves;

Step 2: Given a subset of subcarriers, each user solves the optimization problem of maximizing its own distortion reduction under the power constraint;

We then iteratively update both the subcarrier assignments (according to the application layer distortion) and the power allocation strategy (based on the CSI). One of the major differences between our algorithm and [41] and [42] is that we allow application layer information and physical layer information to interact in our decision process. Before providing the details of the algorithm in Section 3.3, we first investigate a condition for the optimal solution in a continuous channel setting, where there can be variations within a subcarrier, as opposed to a block fading model. This condition inspires our algorithm.
3.2 Continuous Frequency Channel Response Resource Allocation Analysis

The optimization problem defined in Section 3.1 involves both bandwidth partitioning and subcarrier power allocation. Because of the discrete nature of the subcarriers, the resource allocation decisions (bandwidth partitioning of the band $W$) lie in a set of discrete points, and thus it is difficult to analyze as an optimization problem. In this section, we remove the constraint of subcarrier block fading and try to find the optimal allocation decision for a continuous-frequency channel response. We show that for the cross layer design problems defined in the previous section, the optimal condition is jointly dependent on the slope of the RD curve and a physical layer metric which measures the relative bandwidth utilization efficiency.

3.2.1 Conventional Video Multiplexing

We first consider conventional two-user video multiplexing, where a sum of rate $r$ (assuming error free) is shared by two users. Given that the resource pool here is bit rate, the allocator needs to decide a pair of rates $(r_1, r_2)$ for the optimization problem:

$$
\min_{r_1, r_2} \left( a_1 + \frac{b_1}{r_1 + c_1} \right) + \left( a_2 + \frac{b_2}{r_2 + c_2} \right)
$$

subject to $r_1 + r_2 \leq r$

To solve (3.2), Karush-Kuhn-Tucker conditions [59] yield the following four relations for the optimal allocation $(r_1^*, r_2^*)$:

(3.2a) Primal feasibility: $r_1^* + r_2^* - r \leq 0$;

(3.2b) Dual feasibility: $\nu \geq 0$;

(3.2c) Complementary slackness: $\nu(r_1^* + r_2^* - r) = 0$;

(3.2d) Stationarity: $\frac{b_1}{(r_1^* + c_1)^2} = \frac{b_2}{(r_2^* + c_2)^2} = \nu$; where $\nu$ is a Lagrange multiplier.

Note that the value of the Lagrange multiplier is the absolute value of the RD functions at optimal rate allocation $r_1^*$ and $r_2^*$. In [7] and [8], the optimal condi-
tion for this optimization problem is described as the ‘constant slope optimization’ or ‘equal slope condition’.

Alternatively, one can view (3.2d) as a stationary condition (optimal condition) such that any switch of source rate $\Delta r > 0$ among two users would cause nonnegative change of the sum of distortion.

To prove this, if $\left(r^*_1, r^*_2\right)$ is the optimal rate allocation pair, switching rate $\Delta r$ from user 2 to user 1, we will have

$$\frac{b_1}{r^*_1 + c_1} + \frac{b_2}{r^*_2 + c_2} \leq \frac{b_1}{(r^*_1 + \Delta r) + c_1} + \frac{b_2}{(r^*_2 - \Delta r) + c_2}$$

(3.3)

where we drop the $a_k$ terms because they are constant with respect to the rate $r$.

(3.3) can be reduced to

$$\frac{b_1 \Delta r}{(r^*_1 + c_1)^2 + \Delta r(r^*_1 + c_1)} \leq \frac{b_2 \Delta r}{(r^*_2 + c_2)^2 - \Delta r(r^*_2 + c_2)}$$

(3.4)

Since the RD functions are strictly convex, the condition in (3.4) must be satisfied for $\Delta r \to 0$. We then can drop $\Delta r(r^*_1 + c_1)$, and $\Delta r(r^*_2 + c_2)$ as they are infinitesimally small compared to the square terms, and we have

$$\frac{b_1}{(r^*_1 + c_1)^2} \leq \frac{b_2}{(r^*_2 + c_2)^2}$$

(3.5)

Similarly, switching rate from user 1 to user 2 will also cause a nonnegative change of the sum of distortions, and we have

$$\frac{b_1}{(r^*_1 + c_1)^2} \geq \frac{b_2}{(r^*_2 + c_2)^2}$$

(3.6)

Combining (3.5) and (3.6), we reach the equal slope condition,

$$\frac{b_1}{(r^*_1 + c_1)^2} = \frac{b_2}{(r^*_2 + c_2)^2}$$

(3.7)

We now see that the stationarity condition (3.2d) can be proved from a resource switching point of view. In [7] and [8], this idea of resource switching is used to design a source rate allocation algorithm. The algorithms will converge to the optimal equal slope condition.
We will now change the resource from source rate to frequency bandwidth and consider an allocation problem where the source rate is a function of wireless bandwidth and transmission power.

### 3.2.2 Two-User Continuous Frequency Channel Response Optimal Allocation

We now consider a two-user system in a continuous channel setting. In this scenario, the allocator can divide the total frequency band \( W \) into infinitely small bands for resource allocation. We use \( |H_k(f)|^2 \) to denote the channel gain for user \( k \) at frequency \( f \).

Let \( \hat{B}_i \) be the frequency band assigned to user \( i \). If we ignore the upper bound of the modulation alphabet size, the optimization problem becomes:

\[
\min_P \sum_{k=1}^{2} \int_{B_k} b_k \log_2[1 + \eta P_k(f)|H_k(f)|^2] df + c_k
\]

subject to

(C1) \( \hat{B}_1 \cap \hat{B}_2 = \emptyset \), \( (B_1 \cup B_2) \subset W \)

(C2) \( \int_{B_k} P_k(f) df \leq P, k = 1, 2 \)

Here, \( P_k(f) \) should follow the water filling solution after the band allocation is decided. Given \( W \), the optimal band allocation can be viewed as a partition of the band \( W = B_1^{opt} \cup B_2^{opt} \cup B^{extra} \). Here, \( B_1^{opt} \) and \( B_2^{opt} \) are the optimal sets of frequency bands assigned to two users in the sense that the sum of distortions is minimized, and no frequency component in \( B_i^{opt} \) would exceed the water level of user \( i \). \( B^{extra} \) is the set of bands not assigned to either user. We introduce the following definitions.

**Definitions I:**

a) Let \( |\cdot| \) be the bandwidth in Hz, e.g., \( |B_1^{opt}| \) is the optimal bandwidth assigned to user 1.

b) Let \( r_1 = \int_{B_1^{opt}} \log_2(1 + \eta P_1(f)|H_1(f)|^2) df \) and \( r_2 = \int_{B_2^{opt}} \log_2(1 + \eta P_2(f)|H_2(f)|^2) df \) be the average optimal rates (in bits/sec) of two users. Here, \( |H_1(f)|^2 \) and \( |H_2(f)|^2 \)
are the frequency channel responses of the two users. Let $P_1(f)$ and $P_2(f)$ be the power allocations which obey the water filling solution [4].

c) Define $W_1 = P_1(f) + \frac{1}{\eta |H_1(f)|^2}$ and $W_2 = P_2(f) + \frac{1}{\eta |H_2(f)|^2}$ to be the water levels for the two users at the optimal solution (Fig. 3.1 and Fig. 3.2).

d) Let $\theta \in B_2^{opt}$ be an infinitesimally small band $\theta$ assigned to user 2. Note that $|H_1^\theta|^2$ and $|H_2^\theta|^2$ are the channel gains for user 1 and user 2 for band $\theta$, respectively. They are constant since the band is infinitesimally small.

e) Let $\phi_i = \left( W_i - \frac{1}{\eta |H_i(f)|^2} \right)^+$ be the non-negative distance between the water level of user $i$ and the noise level of band $|H_i^\theta|^2$. By definition, $[x]^+ = x$ if $x > 0$ and $[x]^+ = 0$ if $x \leq 0$. For any frequency band $\theta$ of $B_2^{opt}$, $W_2 - \frac{1}{\eta |H_2^\theta|^2} > 0$. For user 2, the value of $\phi_2^\theta$ is always positive.

\[ \begin{align*}
\frac{1}{|H_1^\theta|^2} & \quad \frac{1}{\eta |H_1(f)|^2} \\
W_i & \quad \hat{B}_i \quad f_0 \quad \theta \quad f
\end{align*} \]

\[ P_{1,\theta} \]

**Figure 3.1:** Water level change for user 1 gaining one band. The water level drops from $W_1$ to $W_1'$ after user 1 gains one additional band $\theta$.

**Theorem 1:** For a continuous frequency channel $W$, the optimal band allocation of $B_1^{opt}$ and $B_2^{opt}$ for minimizing the sum of distortions should satisfy (3.9), for any frequency band $\theta$ assigned to user 2.
Figure 3.2: Water level change for user 2 losing one band. The water level rises from $W_2$ to $W'_2$ after user 2 loses one band $\theta$.

\[
\begin{align*}
\frac{b_1}{(r_1+c_1)^2} & \left\{ \ln \left( 1 + \eta \phi_1^\theta |H_1^\theta|^2 \right) - \int_{B^\text{opt}_1} \frac{\eta |H_1(f)|^2}{|B^\text{opt}_1|(1+\eta P_1(f)|H_1(f)|^2)} \phi_1^\theta df \right\} \\
\frac{b_2}{(r_2+c_2)^2} & \left\{ \ln \left( 1 + \eta \phi_2^\theta |H_2^\theta|^2 \right) - \int_{B^\text{opt}_2} \frac{\eta |H_2(f)|^2}{|B^\text{opt}_2|(1+\eta P_2(f)|H_2(f)|^2)} \phi_2^\theta df \right\} \\
\end{align*}
\]

\[\leq 1 \quad (3.9)\]

**Proof:** If an assignment is optimal, any reassignment will not decrease the sum of distortions. Let $B^\text{opt}_1$, $B^\text{opt}_2$ be the optimal assignment, and let $B^\text{opt}_1 \cup \theta$, $B^\text{opt}_2 - \theta$ be a new assignment which reassigns band $\theta$ to user 1. If an assignment is optimal, then

\[
\frac{b_1}{r_1 + c_1} + \frac{b_2}{r_2 + c_2} \leq \frac{b_1}{(r_1 + \Delta r_1) + c_1} + \frac{b_2}{(r_2 - \Delta r_2) + c_2} \quad (3.10)
\]

where $\Delta r_1$ and $\Delta r_2$ are the rate changes caused by switching band $\theta$. We have two scenarios.

**Scenario A:** $W_1 > \frac{1}{\eta |H_1^\theta|^2}$ or $\phi_1^\theta > 0$
In this case, user 1 would get positive rate gain by acquiring the additional band $\theta$. In other words, $\Delta r_1 > 0$ and $\Delta r_2 > 0$. Continuing from (3.10), we can go one step further and get

$$\frac{b_1}{(r_1 + c_1)^2 + \Delta r_1 (r_1 + c_1)} \Delta r_1 \leq \frac{b_2}{(r_2 + c_2)^2 - \Delta r_2 (r_2 + c_2)} \Delta r_2$$

(3.11)

Because $D(R) = a_k + \frac{b_k}{R + c_k}$ is strictly convex, (3.10) must be satisfied as we take $|\theta| \to 0$. It is easy to see that as $|\theta| \to 0$, $\frac{\Delta r_i}{r_i + c_i} \to 0$ for $i=1$ and 2. We thus can drop $\Delta r_i (r_i + c_i)$ terms as they will be infinitely small compared to the squared term. So the optimal condition is:

$$\frac{b_1}{(r_1 + c_1)^2} \Delta r_1 \leq \frac{b_2}{(r_2 + c_2)^2} \Delta r_2$$

(3.12)

Now, we are interested in finding $\lim_{|\theta| \to 0} \frac{\Delta r_1}{\Delta r_2}$, which is the ratio of rate change as $|\theta| \to 0$. Again, in the new frequency assignment, user 1 gets $B_1^{opt} \cup \theta$ and user 2 gets $B_2^{opt} - \theta$. Fig. 3.1 shows the power redistribution after switching band $\theta$. $P_{1,\theta}$ is the total power user 1 will put over band $\theta$ after the reassignment. Since we consider $|\theta| \to 0$, $P_{1,\theta}$ is collected uniformly from $B_1^{opt}$ and redistributed uniformly over band $\theta$. Note that $W_1 = P_1(f) + \frac{1}{\eta |H_1(f)|^2}$ is the water level of user 1 before reallocation, and $W_1'$ is the level after reallocation. We then have:

$$W_1 - \frac{P_{1,\theta}}{|B_1^{opt}|^2} = \frac{1}{\eta |H_1|^2} + \frac{P_{1,\theta}}{|\theta|} = W_1'$$

(3.13)

where $|H_1|^2 = |H_1(f_0 + \frac{|\theta|}{2})|^2$ is the channel response over $\theta$, and $f_0$ is the left limit of $\theta$. To go further, we have:

$$P_{1,\theta} = \left( W_1 - \frac{1}{\eta |H_1|^2} \right) \left( \frac{|B_1^{opt}| |\theta|}{|\theta| + |B_1^{opt}|} \right)$$

(3.14)

Before reallocation, the rate for user 1 is:

$$\int_{B_1^{opt}} \log_2 \left( 1 + \eta P_1(f) |H_1(f)|^2 \right) df$$

(3.15)
After getting $\theta$, the new rate is given by:

$$
\int_{B_1^{\text{opt}}} \log_2 \left( 1 + \eta \left( P_1(f) - \frac{P_{1,\theta}}{|B_1^{\text{opt}}|} \right) |H_1(f)|^2 \right) df + \log_2 \left( 1 + \frac{P_{1,\theta}}{\theta} |H_1^{\theta}|^2 \right)
$$

(3.16)

We then can calculate the rate difference as:

$$
\Delta r_1 = \int_{B_1^{\text{opt}}} \log_2 \left( 1 - \frac{\eta P_{1,\theta} |H_1(f)|^2}{|B_1^{\text{opt}}|(1 + \eta P_1(f) |H_1(f)|^2)} \right) df + \log_2 \left( 1 + \frac{P_{1,\theta}}{\theta} |H_1^{\theta}|^2 \right)
$$

(3.17)

Similar to the setting for user 1, $P_{2,\theta}$ is the power allocation for band $\theta$ and $|H_2^{\theta}|^2$ is the frequency response of user 2 over this band. Fig. 3.2 shows the power redistribution after reallocation for user 2, and we can calculate $P_{2,\theta}$:

$$
W_2 - \frac{1}{\eta |H_2^{\theta}|^2} = \frac{P_{2,\theta}}{\theta}
$$

(3.18)

as well as the absolute value of the rate change given by

$$
\Delta r_2 = |\theta| \log_2 \left( 1 + \frac{P_{2,\theta}}{|\theta|} |H_2^{\theta}|^2 \right) - \int_{B_2^{\text{opt}}} \log_2 \left( 1 + \frac{\eta P_{2,\theta} |H_2(f)|^2}{|B_2^{\text{opt}}|(1 + \eta P_2(f) |H_2(f)|^2)} \right) df
$$

(3.19)

We are interested in finding the ratio between the rate changes of these two users expressed as

$$
\lim_{|\theta| \to 0} \frac{\Delta r_1}{\Delta r_2} = \lim_{|\theta| \to 0} \frac{\int_{B_1^{\text{opt}}} \log_2 \left( 1 - \frac{\eta P_{1,\theta} |H_1(f)|^2}{|B_1^{\text{opt}}|(1 + \eta P_1(f) |H_1(f)|^2)} \right) df + \log_2 \left( 1 + \frac{P_{1,\theta}}{\theta} |H_1^{\theta}|^2 \right)}{|\theta| \log_2 \left( 1 + \frac{P_{2,\theta}}{|\theta|} |H_2^{\theta}|^2 \right) - \int_{B_2^{\text{opt}}} \log_2 \left( 1 + \frac{\eta P_{2,\theta} |H_2(f)|^2}{|B_2^{\text{opt}}|(1 + \eta P_2(f) |H_2(f)|^2)} \right) df}
$$

(3.20)

We then use L’Hopital’s rule and obtain
\[
\lim_{|\theta| \to 0} \frac{\Delta r_1}{\Delta r_2} = \frac{\ln \left(1 + \eta \phi_1^\theta |H_1^\theta|^2\right) - \int_{B_1^{opt}} \frac{\eta |H_1(f)|^2}{|1 + \eta P_1(f)| |H_1(f)|^2} \phi_1^\theta df}{\ln \left(1 + \eta \phi_2^\theta |H_2^\theta|^2\right) - \int_{B_2^{opt}} \frac{\eta |H_2(f)|^2}{|1 + \eta P_2(f)| |H_2(f)|^2} \phi_2^\theta df}
\]

(3.21)

From (3.12) and (3.21), for a two-user uplink video transmission scenario, the optimal frequency and power allocation scheme should satisfy (3.22) for any frequency band \( \theta \) assigned to user 2.

\[
\frac{b_1}{(r_1 + c_1)^2} \left\{ \ln \left(1 + \eta \phi_1^\theta |H_1^\theta|^2\right) - \int_{B_1^{opt}} \frac{\eta |H_1(f)|^2}{|1 + \eta P_1(f)| |H_1(f)|^2} \phi_1^\theta df \right\} 
\leq 1
\]

\[
\frac{b_2}{(r_2 + c_2)^2} \left\{ \ln \left(1 + \eta \phi_2^\theta |H_2^\theta|^2\right) - \int_{B_2^{opt}} \frac{\eta |H_2(f)|^2}{|1 + \eta P_2(f)| |H_2(f)|^2} \phi_2^\theta df \right\}
\leq 1
\]

(3.22)

Scenario B: \( W_1 \leq \frac{1}{\eta |H_1^\theta|^2} \) or \( \phi_1^\theta = 0 \)

This condition means that when we try to switch band \( \theta \) from user 2 to user 1, the frequency response of user 1 over this band does not exceed the original water level, and the optimal solution will not put any power into this band. In this case \( \Delta r_1 = 0 \), \( \Delta r_2 > 0 \), and \( \phi_1^\theta = 0 \). If we plug \( \phi_1^\theta = 0 \) into the numerator of (3.21), we have (3.23)

\[
\frac{b_1}{(r_1 + c_1)^2} \left\{ \ln \left(1 + \eta \phi_1^\theta |H_1^\theta|^2\right) - \int_{B_1^{opt}} \frac{\eta |H_1(f)|^2}{|1 + \eta P_1(f)| |H_1(f)|^2} \phi_1^\theta df \right\} = 0 < 1
\]

(3.23)

Combining both scenarios, for a two-user uplink video transmission scenario, the optimal frequency and power allocation scheme should satisfy (3.22).

Similarly, in a system with an arbitrary number of users, it is easy to conclude that, for frequency band \( \theta \) to be assigned to user \( j \), the following condition must be satisfied for any user \( i \neq j \).
To find the optimal allocation in this cross layer problem, we wish to maximize the combination of application and physical layer metrics, which as seen in (3.9), is the product of the absolute value of application layer RD slope

\[
\frac{b_i}{(r_i + c_i)^2} \left\{ \ln \left( 1 + \eta \phi_i^\theta |H_i^\theta|^2 \right) - \int_{B_{i}^{opt}} \frac{\eta |H_i(f)|^2}{|B_{i}^{opt}| \left( 1 + \eta P_i(f)|H_i(f)|^2 \right) \phi_i^\theta df} \right\} \leq 1 \tag{3.24}
\]

and the physical layer information given by:

\[
\frac{b_j}{(r_j + c_j)^2} \left\{ \ln \left( 1 + \eta \phi_j^\theta |H_j^\theta|^2 \right) - \int_{B_{j}^{opt}} \frac{\eta |H_j(f)|^2}{|B_{j}^{opt}| \left( 1 + \eta P_j(f)|H_j(f)|^2 \right) \phi_j^\theta df} \right\} \leq 1 \tag{3.26}
\]

For video RD characteristics in the form of \( D_i = a_i + \frac{b_i}{r_i + c_i} \), (3.25) is the absolute value of the slope of the RD curve for user \( i \) at rate \( r_i \). In this sense, for an allocation scheme to be optimal, the application layer contribution to the overall metric should be the slope of the curve instead of the distortion value [7], [8]. To solve the optimization problem of (3.1), the algorithm should give priority to the user with the steepest slope. On the other hand, (3.26) is an explicit relation between the physical layer rate (in bits/sec) and channel state information. As the bandwidth of \( \theta \) becomes infinitesimally small, (3.26) can be considered as the marginal rate change (either increase or decrease) of switching a band from one user to the other. More specifically, one may treat \( \ln \left( 1 + \eta \phi_i^\theta |H_i^\theta|^2 \right) \) as the direct rate change caused by gaining or losing \( \theta \), and \( \int_{B_{i}^{opt}} \frac{\eta |H_i(f)|^2}{|B_{i}^{opt}| \left( 1 + \eta P_i(f)|H_i(f)|^2 \right) \phi_i^\theta df} \) as the corresponding rate decrease or increase due to the effect of water level change.

The optimal cross layer allocation would assign band \( \theta \) to the user who has the maximum physical layer marginal rate increase given by (3.26), weighted by the slope of the RD curve.
To solve (3.1), given finite subcarrier bandwidths, the physical layer metric expression of (3.26) would not be valid, as the frequency bands are modeled as experiencing block fading. We thus design an iterative subcarrier allocation algorithm in the next section. Similar to the optimal condition derived in (3.9), the application layer metric is the slope of the RD curve. We will give users with steep slope priority to access subcarriers. In the continuous channel response allocation analysis, the increment considered for switching between users was infinitesimal, whereas in the algorithm, the increment is the bandwidth of a single subcarrier.

### 3.3 Iterative Resource Allocation Algorithm

To find a solution to the problem defined in (3.1), we design an iterative algorithm which allows physical layer CSI and application layer RD information to interact. To describe this algorithm, we use the notation for instantaneous CSI in this section. This algorithm first assigns the subcarriers purely based on channel conditions. However, it is possible that the overall performance (from an average distortion perspective) might be better if we assign some subcarriers to a user with worse channel conditions, but who might need a greater bit rate. We then try to reassign one subcarrier to the user with the steepest distortion curve slope. To solve a conventional video multiplexing bit rate allocation problem, a condition for a global optimum is that users operate at a rate with the same slope of their corresponding RD curves [7], [39], [60]. Note that at each iteration we only change the assignment of one subcarrier through a search process, and for every subcarrier which is not assigned to the user with the steepest slope, the calculations of distortion loss for the user losing that subcarrier and the performance improvement for the user with the steepest slope gaining that subcarrier is of low complexity. We then make the reassignment of the subcarrier that can most effectively reduce the overall distortion. We repeat this procedure iteratively until we run out of the possibility of reassigning subcarriers. We introduce the following definitions that will be used in the algorithm.

**Definitions II:**
a) Let $\rho_m^{(i)}$ denote the user who is assigned subcarrier $m$ at the $i$-th iteration. For example, $\rho_2^{(1)} = 3$ means user 3 is assigned subcarrier 2 at the first iteration of the algorithm.
b) Define $A_k^{(i)}$ to be the set of subcarriers assigned to user $k$ at the $i$-th iteration.
c) Define the potential set $\Omega$ as the set of users that have the potential to improve the average overall system performance by receiving extra subcarriers, and define $|\Omega|$ as the cardinality of the potential set.
d) Define $\Delta_{k,m} \geq 0$ as the absolute value of the video distortion change of user $k$ by gaining or losing subcarrier $m$.

**Iterative Cross Layer Resource Allocation Algorithm:**

**Step (1) Initialization:** Initialize the potential set $\Omega = \{1, 2, 3...K\}$. Initialize $\rho_m^{(0)} = \arg \max_k \{|H_{k,m}|^2\}$ for $m \in \{1, 2, 3...M_c\}$. We first assign each subcarrier to the user who has the best channel response, and let the potential set be the total set.

**Step (2) Water Filling and Slope Calculation:** After subcarrier assignment, each user tries to solve a MSE distortion minimization problem as follows:

$$
\min_{P_{k,m}} \frac{b_k}{\sum_{m \in A_k^{(i)}} \log_2[1 + \eta P_{k,m}|H_{k,m}|^2] + c_k}
$$

$$
\text{s.t. } \sum_{m \in A_k^{(i)}} P_{k,m} \leq P
$$

The optimization problem can be further simplified as

$$
\max_{P_{k,m}} \sum_{m \in A_k^{(i)}} \log_2[1 + \eta P_{k,m}|H_{k,m}|^2]
$$

The solution to this problem is the conventional power water filling allocation [4]:

$$
P_{k,m}^* = \left[ \frac{1}{\lambda_k} - \frac{1}{\eta |H_{k,m}|^2} \right]^+, \forall m \in A_k^{(i)}
$$

where $\lambda_k$ can be found numerically to make the sum of power equal to $P$. Let

$$
r_k^* = \sum_{m \in A_k^{(i)}} \log_2[1 + \eta P_{k,m}^*|H_{k,m}|^2]
$$
be the optimal rate (in bits/symbol) user $k$ gets using water filling. Then $S_k = \left. \frac{d}{dr_k} \frac{b_k}{r_k^p} \right|_{r_k=r_k^*} = -\frac{b_k}{(r_k^* + c_k)^2}$ is the slope of the $k$-th user’s RD curve evaluated at the rate that user $k$ is assigned. Let $k^* = \arg\min_{k \in \Omega} \{S_k\}$ be the user with the steepest slope in the potential set. This is the user who stands to benefit the most from receiving an increment of rate.

Step (3) Subcarrier Reassignment:

We consider taking one subcarrier away from one other user in $\Omega$ and reassigning it to user $k^*$, as user $k^*$ has the largest marginal performance increment in the potential set. We consider each subcarrier $m \in \{1, 2, 3...M_c\} \setminus A^{(i)}_{k^*}$, which is not currently assigned to user $k^*$. We calculate the MSE performance change $-\Delta \rho^{(i)}_{m^*} < 0$ of user $\rho^{(i)}_m$ from losing one subcarrier, and the performance gain of the user $k^*$, $\Delta_{k^*,m} > 0$. The rate change for switching subcarriers is similar to the derivation for (3.26) in a continuous channel case, and the details of the calculation are in the Appendix. Since we only take one subcarrier from one user each time, the MSE performance loss and gain can be found analytically. We then find $m^* = \arg \max_{m \in \{1, 2, 3...M_c\} \setminus A^{(i)}_{k^*}} (\Delta_{k^*,m} - \Delta \rho^{(i)}_{m^*})$, which maximizes the performance change.

If $(\Delta_{k^*,m^*} - \Delta \rho^{(i)}_{m^*}) > 0$, we reassign subcarrier $m^*$ to user $k^*$ at iteration $i+1$, $\rho^{(i+1)}_m = k^*$, and return to Step (2) to update $k^*$.

If $(\Delta_{k^*,m^*} - \Delta \rho^{(i)}_{m^*}) < 0$, which means that the overall performance will not be enhanced by reassigning any subcarrier to user $k^*$, we update the potential set $\Omega = \Omega \setminus \{k^*\}$. User $k^*$ is dropped from the potential set. User $k^*$ will keep the subcarriers already assigned to him but will not be assigned any additional subcarriers. Next, we check the cardinality of $\Omega$. If $|\Omega| = 1$, we stop, otherwise, $i$ is incremented, and we go back to step (2) to update $k^*$.

Based on our analysis in Section 3.2, from the perspective of minimizing the sum of distortions, the subcarrier assignment should balance both the application layer metric (the slope of the RD curve) and the physical layer metric. The initialization step, which is purely based on the physical layer metric, will most likely mismatch the optimal criteria we described in (3.26). The idea of reassignment is that, when we are using the iterative method to allocate limited resources, the
user operating at the steepest rate distortion curve has the priority to be assigned extra subcarriers. A flow chart of the algorithm can be found in Fig. 3.3.

3.4 Baseline Algorithms

We compare the performance of our cross layer optimization algorithm to two baseline algorithms, one with only application layer RD information and the other with only physical layer CSI available for resource allocation at the base station.

3.4.1 Application Layer Optimization Algorithm

The application layer optimization allocates subcarriers purely based on the RD information of the video streams. Since CSI is not used, the allocator will treat all subcarriers the same when making the allocation decision. As we will see in the numerical results, to determine the number of subcarriers assigned to each user, we first choose a PSNR value (e.g., PSNR=28dB). To achieve this PSNR, user $k$ needs video coding rate $r_k$ based on his RD information. The number of subcarriers assigned to the $k$-th user is proportional to the $k$-th user’s rate, or

$$L_k \sim M_c \cdot \frac{r_k}{\sum_i r_i} \quad (3.32)$$

where $M_c$ is the total number of subcarriers in the system. Subcarriers are then randomly assigned to users.

After being informed of the resource allocation decision, we assume that users know their CSI, and can use it to select their modulation format. In other words, CSI is not used for resource allocation, but is used to determine the transmitted waveform. Similar to the cross layer and physical layer optimization algorithms, user $k$ conducts a water filling calculation for transmission power assignment, and the modulation format is chosen based on (2.5) for each subcarrier. The source encoding rate is then determined using (2.7).
Figure 3.3: Uplink optimization algorithm flow chart
3.4.2 Physical Layer Optimization Algorithm

Suppose \([H_{1,m}, H_{2,m}...H_{K,m}]\) is the vector of channel gains of users \(\{1,2...K\}\) for subcarrier \(m\). The conventional resource allocation based on multi-user diversity (MUD) would assign subcarrier \(m\) to the user \(k^*\), where:

\[
k^* = \arg \max_k \left\{ \frac{|H_{k,m}|^2}{|H_k|^2} \right\}
\]

(3.33)

Here, \(|H_k|^2 = \frac{1}{M_c} \sum_{m=1}^{M_c} |H_{k,m}|^2\) is the empirical average channel gain of user \(k\). After subcarrier assignment, every user would apply water filling to allocate power to each assigned subcarrier.

Define \(B_c\) to be the coherence bandwidth of the system. For simplicity, we assume that the coherence bandwidth is always an integer multiple of the subcarrier bandwidth, i.e., \(B_c = \Psi W/M_c, \Psi \in \mathbb{Z}^+\) in the simulation. Further, we assume that the channel gains are identical within the coherence bandwidth, but independent between different coherence bands.

For a system with coherence bandwidth larger than the subcarrier bandwidth, i.e., \(\Psi > 1\), a MUD-based algorithm proposed by [30] [31] allocates subcarriers in chunks, i.e., if a given user is assigned a particular subcarrier, that user will also get all the other subcarriers in the chunk. For a system using MUD with large \(\Psi\), since individual users could get multiple chunks with large bandwidth, the resource allocation might be unbalanced and the average video performance will suffer a large degradation. To avoid a scenario where a few users dominate the use of the subcarriers, we design an algorithm that limits the number of subcarriers assigned to each user.

Definitions III:

a) Define \(\Lambda\) as the set of users who are eligible for being assigned additional subcarriers;

b) Define \(\Theta\) as the set of users who have not been assigned any subcarrier yet in the iteration. We design the algorithm such that each user will get at least one subcarrier;

c) Define \(\Gamma\) as the set of subcarriers whose allocation decision has not been made
yet;
d) Similar to the application layer optimization algorithm, let $L_k$ be the number of subcarriers user $k$ is assigned. To control the degree of imbalance in the number of subcarriers that users receive, we impose a set of thresholds $\psi_n$, $n = 1, 2...K - 1$, such that the sum of subcarriers for any group of $n$ users will not exceed $\psi_n$.

We set $\psi_n$, for $1 \leq n \leq K - 1$, equal to:

$$\psi_n = \psi_{n-1} + \left\lceil \epsilon \left( \frac{M_c - \psi_{n-1}}{K - (n - 1)} \right) \right\rceil$$

(3.34)

where, for $n = 1$, this expression reduces to $\psi_1 = \left\lceil \epsilon \frac{M_c}{K} \right\rceil$. In (3.34), the parameter $\epsilon$ is chosen to be greater than or equal to 1, and controls the imbalance of the resource allocation. A larger value of $\epsilon$ means that the resource allocation decision will be more unbalanced, biased to the users who have larger channel gains. For each individual user, the number of subcarriers threshold $\psi_1$ is set to be $\epsilon$ times larger than the average number of subcarriers per user, $M_c/K$. Assuming that one user has already been assigned the maximum of $\psi_1 = \left\lceil \epsilon \frac{M_c}{K} \right\rceil$ subcarriers, the average number of subcarriers for the remaining $(K - 1)$ users is given by $(M_c - \psi_1)/(K - 1)$ and the resource for any combination of two users is limited by $\psi_1 + \left\lceil \epsilon (M_c - \psi_1)/(K - 1) \right\rceil$ subcarriers. We repeat this process iteratively for $n \leq (K - 1)$, and the total number of subcarriers assigned to any group of $n$ users can be found iteratively using (3.34). As a specific example, consider a system with 1000 subcarriers, 3 users, and $\epsilon = 1.5$. The threshold would be $\psi_1 = 500$ subcarriers for each individual user, and $\psi_2 = 875$ subcarriers for any group of two users. When the coherence bandwidth is equal to the entire bandwidth, the user with the strongest channel gain will get 500 subcarriers. Any group of two users cannot get more than 875 subcarriers, so the user with second best channel gain gets 375 subcarriers. The remaining 125 subcarriers are assigned to the third user. When the coherence bandwidth becomes smaller, it will be increasingly unlikely that the total number of subcarriers for a group of $n$ users will reach the threshold of $\psi_n$.

**Physical Layer Optimization Algorithm:**

**Step 1 Initialization:** We initialize $\Lambda$ and $\Theta$ as the complete set of users,
i.e. \( \Lambda = \{1, 2, \ldots, K\} \), \( \Theta = \{1, 2, \ldots, K\} \), \( \Gamma \) as the complete set of subcarriers \( \Gamma = \{1, 2, \ldots, M_c\} \) and \( \psi_n \) as (3.34).

**Step 2 Subcarrier Assignment**: We choose the best channel gain from all the possible assignments,

\[
(k^*, m^*) = \arg \max_{k \in \Lambda, m \in \Gamma} \left\{ \frac{|H_{k,m}|^2}{|H_k|^2} \right\}
\]  

(3.35)

with \( |H_k|^2 = \frac{1}{M_c} \sum_{m=1}^{M_c} |H_{k,m}|^2 \), and assign subcarrier \( m^* \) to user \( k^* \). We update \( \Gamma = \Gamma \setminus m^* \). If \( k^* \in \Theta \), we update \( \Theta = \Theta \setminus k^* \), meaning that user \( k^* \) has been assigned at least one subcarrier. Here, \( |H_{k^*,m^*}|^2 \) stands for the best channel response in all possible subcarrier assignment combinations at the current step.

**Step 3 Status Update**: We check the remaining resource and conduct the following two updates:

1) For every \( n \), \( 1 \leq n \leq K - 1 \), we compare the sum of subcarriers for all groups of \( n \) users with \( \psi_n \). If the sum is equal to \( \psi_n \) for any group, all the users in that group will be excluded from \( \Lambda \).

2) We then check the relation between the number of subcarriers left and the cardinality of \( \Theta \). To ensure that each user can get at least one subcarrier, if \( |\Gamma| = |\Theta| \), we will terminate the algorithm by assigning exactly one of the remaining unallocated subcarriers to each of the users who has no subcarrier yet using (3.35).

We then go back to Step 2 and repeat (3.35) to assign subcarriers until \( \Gamma \) is empty.

### 3.5 Simulation Setup

#### 3.5.1 Physical Layer Setup

We study an uplink multi-carrier system with 16 subcarriers, each with a bandwidth of 50 kHz. We evaluate performance by the Peak-Signal-to-Noise Ratio (PSNR), defined as: \( PSNR = 10\log_{10} \frac{255 \times 255}{MSE} \). For all three optimization algorithms, the modulation decision will be rounded down to a valid integer value corresponding to a modulation format of MQAM, with \( M=4, 8, 16, 32, 64, 128 \) or...
256 based on (2.5). For example, if the cross layer allocation assigns a rate of any real value \( R_{1,4} \in [3, 4] \) for user 1 on subcarrier 4, the actual alphabet size would be 8-QAM. The channel response consists of both the path loss and multi-path fading, and the magnitude of the channel can be written as \( |H_{k,m}|^2 = \alpha^2 \cdot K_0 \cdot \left( \frac{d_0}{d_k} \right)^\gamma \) [4], where \( \gamma = 2.4 \) is the path-loss exponent [4]. \( d_k \) is the distance between user \( k \) and the base station, and \( d_0 \) is a reference distance set to 10m [4]. In addition, \( \alpha \) is assumed to be a Rayleigh random variable, and \( K_0 \) is a constant of -24dB. We assume that the distance \( d_k \) between user \( k \) and the base station is a random variable, and follows a uniform distribution between [30, 120] meters. For a user who is 75 meters away from the base station, the average SNR is assumed to be 18 dB if only one subcarrier is assigned. Unless otherwise stated, the subcarriers are assumed to fade independently. For the physical layer optimization algorithm, we set \( \epsilon = 1.5 \), which means that one user cannot be assigned more than 150% of the average number of subcarriers.

For all three optimization schemes, we use a rate 1/2 convolutional code with code generator polynomial [23, 35] in octal, and the coded bits are interleaved across different subcarriers. For example, if one user gets three subcarriers, the first coded bit goes to the first subcarrier, the second coded bit goes to the second subcarrier, etc. We use log-likelihood ratio demodulation to detect each bit of the QAM symbol. We then decode the bitstream using soft-decision decoding with eight reliability ranges.

### 3.5.2 Application Layer Setup

We use a sequence of CIF videos of total length 50 seconds at 30 frames per second. Compression is by the baseline profile of H.264/AVC reference software JM 11.0 [54]. The GOP size is 15 frames (I-P-P-P) and the frames inside one GOP are encoded using H.264 rate control. We encode each GOP at rates of 80, 100, 120, 140, 160, 180, 200, 220, 240, 280, 300, 340, 380, 420, 460, 500 and 600 kbps, and use these operational points to fit the rate distortion function \( D(R) = a_k + b_k/(R+c_k) \) by nonlinear regression. We randomly assign different starting points of the same video to different users, and the resource allocation decision is done at the start of
every GOP. The video in the simulation is a travel documentary which consists of both high motion and low motion GOPs. By assigning random starting points of the same cyclic video to different users, we create application layer diversity among users and yet have the same average complexity over time for different users. Each video is encoded at 10 slices per frame, and any channel error will make the system lose the entire slice. At the decoder side, slice copying conceals losses.

3.6 Results

3.6.1 Systems with Different $SER_t$

As discussed in Chapter 2, the uplink resource allocation algorithm needs an input $SER_t$ value; we used $SER_t = 0.2$, and varied the number of users from 4 to 12 in the system of 16 subcarriers. Fig. 3.4 shows the performance of the three optimization algorithms. The solid lines represent the numerical results obtained from the RD curves. That is, the resource allocator decides the rate for each user, and the distortion is calculated directly from $D(R)$. This can be considered the error-free distortion, or distortion at the encoder side. The dashed lines are the distortion results measured at the decoder; the videos are reconstructed from the bitstream corrupted by the channel. The effects of packet loss, errors in RD curve fitting, and imperfection of encoder rate control are included in the simulation.

With $SER_t$ set to 0.2, we find that the decoded bit error rate is small, and distortion curves at the encoder and decoder are close. Comparing the performance of these three algorithms, we see that when the number of users in the system is small, the physical layer optimization outperforms the application layer optimization algorithm, and the gap between the cross layer and the physical layer algorithms is relatively small. When the system has abundant resources so each user can be assigned several subcarriers, both cross layer and physical layer optimization algorithms will allow users to operate at a high data rate, or in the flat region of the convex RD curves. Utilizing application RD information in the resource allocation will thus not benefit the overall users’ performance by much.
Figure 3.4: Video PSNR Performance vs. Number of Users. SER_t = 0.2, 16 Subcarriers, Ψ = 1, Average SNR=18dB if only one subcarrier is assigned, Users for PHY Layer Optimization are limited to be assigned at most 1.5 times of the average remaining resources.

Fig. 3.5 shows a sample of the performance for individual users in systems with different numbers of users using the cross layer algorithm. In the first row of the plot, we see that all four users are operating near the right end of their RD curves and the slopes of users are relatively small. When the average resource for each user gets smaller, the users are forced to operate at steeper parts of the RD curves (see the second and third rows of Fig. 3.5). As we increase the number of users in the system to 8 and 12, the gap between the cross layer and the physical layer algorithms widens. We conjecture this is because source characteristics play a more important role when many users compete for the available resources. For a system with many users, it becomes important to combine the information of CSI and RD in the system design for a resource-scarce system, as most users operate on the
steep slope of their individual RD curve. Mismatch of the physical layer resource with the RD curve would cause a large loss of system performance.

**Figure 3.5:** Individual user’s performance in systems with different numbers of users. Each column indicates the same user’s RD relations in systems with four, eight and twelve users.

When the system has 12 users, cross layer optimization outperforms physical layer optimization by about 1.25 dB, and the gap to application layer optimization is even larger. For a system with average PSNR of 30.5 dB, the cross layer scheme can support 12 users, compared to 8 users for physical layer optimization and less than 5 users for application layer optimization. In this sense, the
cross layer algorithm can increase the capacity (the number of users a system can support) by 50% or more.

We now change the value of $SER_t$ to 0.1 and 0.25 (Fig. 3.6 and 3.7, respectively). When we set $SER_t = 0.1$, the modulation alphabet size will be chosen more conservatively and thus force the video source encoding rate to be smaller than for $SER_t = 0.2$. On the other hand, a high $SER_t$ value will lead to a relatively large gap between the encoder curves and decoder curves for PSNR performance at the decoder side, and we see that for $SER_t = 0.25$, the impact of channel errors has significantly decreased the throughput of the system and the PSNR of the video from the error free scenarios. In Fig. 3.8, we compare the video performance of the cross layer algorithm at the encoder and decoder sides for a four-user system. While the PSNR at the encoder monotonically improves as we increase the value of $SER_t$, the gap between the encoder curve and decoder curve widens, and the system with $SER_t = 0.2$ has the best PSNR performance at the decoder side.

Figure 3.6: Video PSNR Performance vs. Number of Users. $SER_t = 0.1, 16$ Subcarriers, $\Psi = 1$, Average SNR=18dB if only one subcarrier is assigned.
Figure 3.7: Video PSNR Performance vs. Number of Users. $SER_t = 0.25$, 16 Subcarriers, $Ψ=1$, Average $SNR=18dB$ if only one subcarrier is assigned

For systems with different $SER_t$, comparing the performance of the three algorithms, we see a similar performance gain of adopting cross layer optimization, and the capacity gain by adopting the cross layer algorithm is still around 1.5.

3.6.2 Systems with Different Coherence Bandwidth

In Fig. 3.9, we set $Ψ = 2$. For simplicity in the simulation, we assume that two adjacent subcarriers have the same realization, and the correlation coefficient between different coherence bands is zero. We observe a very slight performance degradation for both the cross layer and the application layer optimization algorithms. As shown in (3.25) and (3.26), since the cross layer optimization algorithm exploits both physical layer multiuser channel diversity and application layer RD diversity, increasing the coherence bandwidth will not affect the cross layer optimization’s ability to utilize the application layer diversity. Similarly, for application layer optimization, increasing the coherence bandwidth will not change the number
of subcarriers assigned to each user, and the performance loss is very limited. On the other hand, compared to the scenario of $\Psi = 1$, we see a large performance degradation for the physical layer optimization (see Fig. 3.10). As subcarriers will have the same fading realization in groups of two, we lose half of the frequency diversity. Since physical layer optimization does not exploit any application layer diversity, losing frequency diversity at the physical layer will have a big impact on the system performance. If we further increase $\Psi$ to four, as shown in Fig. 3.11 and Fig. 3.12, the performance for the physical layer optimization will further degrade, while the performance of the proposed cross layer algorithm is still robust. Comparing the performance between different algorithms, we see that the cross layer optimization can support 12 users with an average PSNR of about 29.5dB, but the baseline algorithms can at support most 7 users.

Figure 3.8: Video PSNR Performance vs. $SER_t$. 4 Users, 16 Subcarriers, $\Psi=1$, Average SNR=18dB if only one subcarrier is assigned.
3.6.3 Complexity of Iterative Water Filling Algorithm

To show the path of performance improvement from initialization to convergence of the cross layer algorithm, in Fig. 3.13 we plot the average MSE for systems with different numbers of users versus the iteration number. To obtain this plot, we observe the MSE values after each iteration for each individual user and average over the entire video sequence and users. The iteration number equal to one corresponds to the performance of the initialization step. Because of the greediness of the algorithm, the biggest performance improvement occurs in the first few iterations, and MSE curves appear to be concave. After the sixth step, we see a very small performance improvement. As shown in the Appendix, since we can find the performance improvement from switching subcarriers at each iteration analytically, the overall complexity of the proposed algorithm is much lower than that of an exhaustive search.
Figure 3.10: Video PSNR performance degradation comparison. $SER_t = 0.2$, $16$ Subcarriers, $Ψ=1$ vs. $Ψ=2$, Average SNR=18dB if only one subcarrier is assigned.

3.7 Conclusion

In this chapter, we have been focusing on resource allocation for a system with $f_{nd} = 0$. We proposed a cross layer resource allocation framework for transmitting video in an uplink multi-carrier setting, and derived an optimality condition for the bandwidth allocation in a continuous frequency response channel. The power allocation and subcarrier assignment strategy are jointly decided by each user’s CSI as well as RD characteristics. Our analytical results show that the optimal allocation is achieved only if the product of the RD slope and a physical layer metric related to the water filling solution, given by (3.26), is minimized for each band and each user. With a similar technique of switching bandwidth increments as in the analysis, we designed an iterative resource allocation algorithm. At each iteration, our algorithm first evaluates the application layer metric defined by (3.25), and then greedily updates the resource allocation decision jointly
Figure 3.11: Video PSNR Performance vs. Number of Users. $SER_t = 0.2$, 16 Subcarriers, $\Psi = 4$, Average SNR=18dB if only one subcarrier is assigned.

according to (3.25) and (3.26). Compared to a resource allocation using either only application layer or only physical layer information, for the same video performance, the cross layer optimization significantly increased the capacity of the system, and resulted in robust performance as the coherence bandwidth changed, over the range of parameter values considered in our numerical results.

Chapter 3, in part, is a reprint of the material as it appears in Dawei Wang, Laura Toni, Pamela Cosman and Laurence Milstein: "Uplink Resource Management for Multiuser OFDM Video Transmission Systems: Analysis and Algorithm Design", IEEE Transaction on Communications, Vol. 61, No. 5, pp. 2060-2073. The dissertation author was the primary investigator and author of this paper.
Figure 3.12: Video PSNR Performance Degradation Comparison. $SER_t = 0.2$, 16 Subcarriers, $\Psi = 1$ vs. $\Psi = 4$, Average SNR=18dB if only one subcarrier is assigned.
Figure 3.13: *Average MSE vs. Number of Iterations.*
Chapter 4

Resource Allocation For Channels With Arbitrary User Mobility

The problem of cross layer resource allocation can be divided into two categories, based on the relationship between the dynamics of the channel and the time duration over which the video distortion is calculated. In the previous chapter, we studied a scenario for which the channel is constant over the duration of a video GOP. Since we only need to estimate the state of the channel once per resource allocation decision, we can achieve high estimation accuracy with small estimation overhead.

In this chapter, we investigate the problem of multiple video user resource allocation for a system with arbitrary channel variation. The RD function is still measured for GOP duration. When the quality of the channel varies rapidly within the duration of the GOP, the amount of overhead for the estimation of the channel becomes significant, and the tradeoff between the estimation overhead and the channel outing becomes the key in determining the resource allocation performance. Also, since the quality of the channel is not a constant for the duration of GOP, the allocator might not be able to predict the throughput performance of each user when making the resource allocation decision.

Two common approaches can be used in this scenario. To better exploit the multiuser diversity, one can track the dynamics of the channel and make an
allocation decision adaptively every time the state of the channel changes. This method needs extensive information exchange between the users and the allocator, and the computation complexity can be very high. A more realistic approach is to only make one resource allocation for the duration of the GOP, and dynamically adapt the usage of the channel according to the channel quality. This method needs a reasonable amount of information exchange, and the computational complexity is relatively low. We use the second approach, and the resource allocation decision is made on a GOP to GOP basis. Since the channel might experience different fades within the duration of a GOP, the channel throughput cannot be predicted accurately at the beginning of the GOP when the allocator makes the resource allocation decision. After the resource allocation, the users are allowed to adaptively change the modulation format according to the instantaneous channel state.

4.1 Doubly Selective Channel Model

Unlike Chapter 3, we consider a time-varying channel \( f_{nd} > 0 \) with \( M_c \) equally spaced subcarriers over a total bandwidth of \( W \) Hz. The system operates in a slotted manner, and the length of one time slot is \( T_s \) (sec), equal to both the video display time and the transmission duration of one GOP. We assume a block fading model in the frequency domain with coherence bandwidth \( B_c = W/M_c \). Let \( H_{k,m}[l] \) be the complex channel gain of user \( k \) for subcarrier \( m \) at the \( l \)-th symbol. The subcarrier assignment as well as the power allocation decision will be made on a slot-to-slot basis. Each subcarrier can only be used by one user, but it is possible for one user to get more than one subcarrier. Further, \( H_{k,m}[l] = \gamma \alpha_{k,m}[l] \), where \( \gamma \) depends on the path-loss coefficient, the distance between the mobile user and the base station, and the shadowing caused by obstacles. The variable \( \alpha_{k,m}[l] \) captures the multipath fading and is modeled as a zero-mean complex stationary Gaussian random process [49]. The magnitude of \( \alpha_{k,m}[l] \) is Rayleigh distributed with a variance of 1 for a non-line-of-sight (NLOS) system. The band-limited
spectrum of $\alpha_{k,m}(t)$ is given by

$$S(f) = S(0) \left[ 1 - \frac{f}{f_d} \right]^{-1/2}, |f| < f_d$$

(4.1)

Here, $f_d$ is the Doppler spread, and $S(0) = 2/\pi$. Let $f_{nd} = f_d T_{\text{sym}}$ be the normalized Doppler spread, with $T_{\text{sym}}$ as the symbol duration. The autocorrelation between two samples $l$ and $l + \Delta l$ can be written as

$$\mathbb{E} [\alpha_{k,m}[l]\alpha^*_{k,m}[l + \Delta l]] = J_0 [2\pi \cdot f_{nd} \cdot (\Delta l)]$$

(4.2)

where $J_0(\cdot)$ is the zeroth order Bessel function of the first kind. The first zero crossing of the correlation function occurs when the product $f_{nd} \cdot (\Delta l)$ is about 0.4.

### 4.2 Pilot Symbol Assisted Modulation and Multiple User Resource Allocation

In a multiple user system, to minimize the sum of the MSEs across all users, the base station collects the RD information (coefficients of $a_k$, $w_k$ and $v_k$ in (2.6)) as well as the CSI of the subcarriers, and allocates the subcarriers jointly according to application layer and physical layer information. In Chapter 3, we studied video resource allocation for a system in which the channel varies slowly at the GOP level. Under the condition of constant CSI for the entire GOP, we assumed that the modulation format remains unchanged for one GOP, and the throughput of each subcarrier could be perfectly estimated at the beginning of each GOP. The resource allocation problem then becomes a mixed integer programming problem, and allocation decisions are made on a GOP-by-GOP basis.

#### 4.2.1 System Operation Overview

For mobile users operating in a high enough Doppler environment, the CSI estimated at the beginning of the GOP will be outdated prior to the end of the GOP. For subcarrier $m$, if the modulation format determined by the CSI at the beginning of the GOP is held constant for the entire GOP duration, it is likely
that the video data will either be over-protected or under-protected. On the other hand, the resource allocation decision based on both RD information and CSI is normally of high complexity [41] [42]. If we update the resource allocation decision every coherence time, the base station will need to collect both the instantaneous CSI of all the subcarriers for all the users, as well as the amount of the GOP that has already been transmitted. If the allocator makes a resource allocation decision at each coherence time, it is not only computationally difficult, but also requires a large amount of information exchange. We thus propose a scheme having two phases which balances the computational complexity with the adaptation accuracy.

**Phase I Cross Layer Resource Allocation:** At the beginning of each GOP, user \( k \) submits the RD function \( D_k(B) \) of the current GOP to the base station. The base station measures the instantaneous CSI of the first symbol, \( H_{k,m}[1], (k \in \{1, 2...K\}, m \in \{1, 2..M_c\}) \) of every subcarrier for each user and uses the CSI jointly with the RD information to make an allocation decision. The allocation decision is fed back to the users and each user is allowed to access the subcarriers assigned to him for the entire GOP duration. The resource allocation algorithm and the information exchange for the RD function is conducted once per GOP.

**Phase II Pilot Assisted Adaptive Modulation:** After resource allocation, when each individual user knows the subset of assigned subcarriers, each user periodically sends a pilot symbol to the base station for channel estimation purposes. Based on the instantaneous CSI at each period, the base station updates the modulation format of each subcarrier and feeds the new modulation formats back to the corresponding users. The estimated CSI is also used for demodulation purposes. Since the modulation format is updated periodically, the number of information bits transmitted cannot be estimated accurately at the beginning of the GOP. The important bits will be transmitted first, and the actual number of bits transmitted is determined by the overall channel conditions over the duration of a GOP (\( T_s \) seconds). Note that in this phase, no information about the RD is exchanged, and the resource allocation decision is not updated. The details of the adaptive modulation scheme are discussed below.
4.2.2 Pilot Symbol Assisted Modulation (PSAM)

As depicted in Fig. 4.1, at most one pilot symbol will be sent from the user to the base station in each subcarrier every $L_s$ symbols for channel estimation. Define a time epoch as a group of $L_s$ ($L_s = T_m/T_0$) symbols, which is also the distance between two pilot symbols. The modulation format for every time epoch will be kept the same, and is determined by the CSI of the pilot symbol. We define $\lambda$ as the fraction of information data. In most of this paper, we will focus on a system with independently faded subcarriers, i.e. $\Psi = 1$, we have $\lambda = (L_s - 1)/L_s$.

In a system with coherence bandwidth $\Psi > 1$, since all the subcarriers within the coherence bandwidth will have the same fade realization, fewer pilot symbols need to be sent. For example, in Fig. 4.1, two subcarriers share the same fade ($\Psi = 2$), and one pilot symbol can be saved in each time epoch for every group of two subcarriers spanned by the same coherence bandwidth.

![Figure 4.1: Pilot assisted modulation. One pilot symbol is added for each subcarrier to estimate the channel for every time epoch. If the number of correlated subcarriers is larger than 1, only one pilot symbol is needed for every correlated band.](image)

Let $P_{k,m}$ be the average power for the $k$-th user on the $m$-th subcarrier, where $P_{k,m} = P > 0$ if the $m$-th subcarrier is allocated to the $k$-th user, and $P_{k,m} = 0$ otherwise. Let $\mu_{k,m}$ be the ratio between the power of the pilot symbols
and average data power for the $k$-th user. The power of the pilot symbol, $P^p_{k,m}$, and the data power, $P^d_{k,m}$, are then given by

$$P^d_{k,m} = \frac{P_{k,m} L_s}{L_s - 1 + \mu_{k,m}}$$  \hspace{1cm} (4.3)

$$P^p_{k,m} = \frac{P_{k,m} L_s \mu_{k,m}}{L_s - 1 + \mu_{k,m}}$$  \hspace{1cm} (4.4)

We assume that the value of $\mu_{k,m}$ is the same for all subcarriers of all users, and hence we drop the indices $k$ and $m$. The average power is given by

$$P_{k,m} = \mu \frac{1}{L_s} P^p_{k,m} + \frac{L_s - 1}{L_s} P^d_{k,m}$$  \hspace{1cm} (4.5)

Similar to the previous chapter, we assume flat fading for each subcarrier at the receiver, and the $l$-th output signal of user $k$ on subcarrier $m$, $Y_{k,m}[l]$, is given by

$$Y_{k,m}[l] = \sqrt{P_{k,m} H_{k,m}[l]} X_{k,m}[l] + N_{k,m}[l]$$  \hspace{1cm} (4.6)

where $x$ is either $p$ or $d$ depending on whether the signal is a pilot symbol or a data symbol, respectively. $X_{k,m}[l]$ is an M-QAM modulated complex signal with unit variance. $N_{k,m}[l]$ is a complex Gaussian random variable with zero-mean and variance $\mathbb{E}\{|N_{k,m}[l]|^2\} = P_N$.

From [50], the symbol error rate (SER) of M-QAM for an AWGN channel can be approximated as

$$SER \approx 4Q \left( \sqrt{\frac{3}{M - 1} \frac{P_{k,m}^p |H_{k,m}[l]|^2}{P_N}} \right)$$  \hspace{1cm} (4.7)

For a given $SER_t$, the modulation format for subcarrier $m$ of user $k$ is updated every $L_s$ symbols based on $\tilde{H}_{k,m}[i L_s + 1]$, which is the estimate of the channel response of the $i$-th pilot, $H_{k,m}[i L_s + 1]$. The information rate (number of bits each symbol can carry in bits/symbol) for the $i$-th group of $L_s$ symbols can be written as $R_{k,m}(P^d_{k,m}, \tilde{H}_{k,m}[i L_s + 1])$ using (2.5):

$$R_{k,m}(P^d_{k,m}, \tilde{H}_{k,m}[i L_s + 1]) = \min\{\lfloor \log_2 \left[ 1 + \eta P^d_{k,m} |\tilde{H}_{k,m}[i L_s + 1]|^2 \right] \rfloor, R_{max}\}$$  \hspace{1cm} (4.8)

where $\eta = \frac{3}{P_N} (Q^{-1} (SER_t/4))^2$ and $R_{max}$ is the largest alphabet size the system allows.
To estimate the channel response $H_{k,m}[iL_s + 1]$, a Wiener filter with $K_e$ pilots is used for interpolation. Since the decision of the modulation format needs to be fed back to the users immediately after the pilot symbol is sent, we can only use the pilots prior to the current one for channel estimation. In other words, to estimate $H_{k,m}[iL_s + 1]$, the pilots at indices $l = jL_s + 1$, $\{j = (i, i-1...-(i-K_e+1))\}$ are used, where $K_e$ is chosen to be even. To estimate the channel gain for the data symbols $H_{k,m}[iL_s + u]$, $u = (2, 3...L_s)$, pilots from both sides can be used. That is, the pilots at time indices $l = jL_s + 1$, $\{j = ((i-K_e/2+1), (i-K_e/2+2)...(i+K_e/2))\}$ are jointly used to interpolate the channel gain.

From [45] [61], the channel estimation error $e = \tilde{H}_{k,m}[l] - H_{k,m}[l]$ can be modeled as a Gaussian random variable with zero mean and variance equal to

$$\sigma_e^2 = \sigma_u^2 - w(l)^{+}R^{-1}w(l)/(P^p_{k,m}|H_{k,m}|^2) \quad (4.9)$$

where $\sigma_u^2 = P^x_{k,m}|H_{k,m}|^2$ is the average received power of the data/pilot symbols, and $+$ is conjugate transpose. If we use $K_e$ pilots for interpolation, $R$ is a $K_e \times K_e$ matrix with the entry in the $i$-th row and $j$-th column given by

$$R_{ij} = P_N\delta_{ij} + P^p_{k,m}|H_{k,m}|^2J_0(2\pi f_{nd}(i-j)L_s) \quad (4.10)$$

where $\delta$ is the Kronecker delta, and $w(l)$ is a $K_e$ column vector for the $l$-th channel sample. The $v$-th row of $w(l)$ is given by

$$w_v(l) = P^p_{k,m}|H_{k,m}|^2\rho_v(l) \quad (4.11)$$

where $\rho_v(l)$ is the correlation coefficient of the $l$-th channel sample and the channel estimate obtained from the $v$-th pilot. For example, as discussed above, to estimate the channel sample $H_{k,m}[l]$, if the symbol at time $l$ is a data symbol and belongs to the $i$-th pilot time epoch, i.e., $iL_s + 2 \leq l \leq (i+1)L_s$, the time epoch of the $v$-th pilot used for interpolation is at $(i-K_e/2+v)L_s + 1$, so that

$$\rho_v(l) = J_0(2\pi f_{nd}[l - ((i-K_e/2+v)L_s + 1)]) \quad (4.12)$$

From (4.9), (4.10) and (4.11), we see that one of the crucial parameters for deciding the channel estimation error is the distance between pilots, $L_s$. For a given
Doppler spread, a smaller $L_s$ results in a larger overhead, and a larger $L_s$ results in larger channel mismatch due to CSI outdating, but increase in throughput. The variance of $e$ is jointly dependent on the number of pilot symbols for interpolation $K_e$ and the ratio $\mu$ between the pilot symbol power and data symbol power.

### 4.2.3 Resource Allocation Problem Formulation

At the beginning of a GOP, the base station estimates the throughput of each subcarrier using (2.5), assuming that the adaptive QAM format will last for $T_s$. The number of bits transmitted over subcarrier $m$ of user $k$ can then be written as $R_{k,m}(P_{k,m}^d, \tilde{H}_{k,m}[1]) \cdot T_s/T_0$, where $T_s/T_0$ is the number of QAM symbols for a GOP. We denote by $\lambda$ ($0 < \lambda \leq 1$) the fraction of data symbols, and $\lfloor (1-\lambda) \cdot T_s/T_0 \rfloor$ symbols will be pilots. To protect the data, a channel code of fixed rate $u$ is added. If the channel stays constant, the number of information bits that the physical layer can support for user $k$ across all $M_c$ subcarriers is given by

$$B_k = \lfloor \sum_{m=1}^{M_c} u \cdot \lambda \cdot R_{k,m}(P_{k,m}^d, \tilde{H}_{k,m}[1]) \cdot T_s/T_0 \rfloor$$

(4.13)

with $P_{k,m}^d = P_d^d$, if subcarrier $m$ is assigned to user $k$, and $P_{k,m}^d = 0$ otherwise.

Although to obtain the performance of the resource allocation algorithm, we will include the effects of channel errors and time varying modulation choices, for the allocation algorithm design, we ignore the effect of channel errors and assume that the modulation format is constant for the GOP duration. We use (4.13) as the channel throughput for our algorithm design. If we plug (4.13) into (2.6), then the MSE distortion for user $k$ can be written as

$$D_k = a_k + \frac{b_k}{\sum_{m=1}^{M_c} \lambda R_{k,m}(P_{k,m}^d, \tilde{H}_{k,m}[1]) + c_k}$$

(4.14)

Here, we have divided both the numerator and denominator by $u \cdot T_s/T_0$ for simplicity. So

$$b_k = \frac{w_k}{(u \cdot T_s/T_0)} \quad c_k = \frac{v_k}{(v \cdot T_s/T_0)}$$

(4.15)
The base station needs to assign $M_c$ subcarriers to $K$ users at the beginning of each GOP, and users can access the subcarrier for the duration of the GOP. The allocation decision will be updated at the beginning of the next GOP as both CSI and RD are updated. Mathematically, our resource allocation goal is to minimize the sum of distortions among $K$ users at each time slot. The optimization objective is

$$\min \sum_{k=1}^{K} \sum_{m=1}^{M_c} \lambda R_{k,m}(P_{k,m}^d, \tilde{H}_{k,m}[1]) + c_k$$

where the entry in the $k$-th row and $m$-th column, $P_{k,m}^d$, is the power allocation of the $m$-th subcarrier for user $k$. The base station sends the allocation decision to the users. We drop the $a_k$ term, as it is constant with respect to $P$. We assume that any subcarrier is used by one user exclusively, so the feasibility constraint for the optimization problem is

(C1) For $m \in \{1, 2, 3...M_c\}$, $P_{k',m}^d P_{k,m}^d = 0$ and $P_{k,m}^d = \{0, P_d\}$

Mathematically, (4.16) is an NP-hard integer programming problem, and an exhaustive search approach would need $K^{M_c}$ calculations. In the next section, we will propose a sub-optimal algorithm which gives priority to users with steep RD curvatures to access the subcarriers. We compare the performance of our algorithm with two baseline algorithms, each of which has limited information about the state of the channel and the state of the videos. Note that, in (4.13), the throughput of each subcarrier is estimated based on the instantaneous CSI at the beginning of the GOP, and the estimation accuracy depends on coherence time.

### 4.3 Resource Allocation Algorithms

Before we state the cross layer algorithm that solves the optimization problem (4.16), we first introduce two baseline algorithms, each of which uses only one layer of information when making the resource allocation decision.
4.3.1 Application Layer Resource Allocation Algorithm

We assume that the base station only knows the application layer information (RD function) when making the allocation decision. Lacking the physical layer CSI, the base station will treat all the subcarriers the same. The resource allocation decision will only specify the number of subcarriers each user is assigned. Define $L_k, (k = \{1, 2...K\})$ as the number of subcarriers allocated to user $k$, which is determined by the relative complexity of the RD functions $D_k(B)$.

**Application Layer Optimization Algorithm:**

**Step 1:** To measure the complexity of each RD curve, we first set a target MSE value $D_t$ (the value is chosen such that the quality of the video is decent, say $D_t = 50$, which corresponds to $PSNR = 31dB$). To achieve this target MSE value, the rate required for user $k$ is given by

$$r_k = \frac{b_k}{D_t - a_k} - c_k$$  \hspace{1cm} (4.17)

We then use $r_k$ as a measure of the complexity of the video for resource allocation purposes.

**Step 2:** Given the constraint of $\sum_{k=1}^{N} L_k = M_c$, the subcarriers will be split such that the number of subcarriers assigned to user $k$ is proportional to $r_k$, or

$$L_k \approx M_c \frac{r_k}{\sum_{k=1}^{K} r_k}$$  \hspace{1cm} (4.18)

$L_k \in \mathbb{Z}$. We use the approximation of (4.18) so that the constraint of $\sum_{k=1}^{K} L_k = M_c$ will be satisfied. To achieve this, we first calculate $\hat{L}_k = M_c \frac{r_k}{\sum_{k=1}^{K} r_k}$ for user $k$, where $\hat{L}_k$ can be any rational number. We then round each $\hat{L}_k$ to $\tilde{L}_k \in \mathbb{Z}$. If $\sum_{k=1}^{K} \tilde{L}_k = M_c$, we have $L_k = \tilde{L}_k, \forall k$. If $\sum_{k=1}^{K} \tilde{L}_k > M_c$, some of the $\hat{L}_k$’s which have been rounded up need to be rounded down. Let $j \triangleq \sum_{k=1}^{K} \tilde{L}_k - M_c$, where $j$ is an integer greater than zero. We choose those $j \hat{L}_k$’s, which have been rounded up and which are farthest
from their respective integer ceilings, and round them down so that the sum of subcarriers across users equals $M_c$. A similar mechanism applies if $\sum_{k=1}^{K} \tilde{L}_k < M_c$.

Since the CSI is not used in the allocation decision, the base station randomly chooses $L_k$ subcarriers for user $k$. After the allocation, the user applies Phase II discussed in the previous section to update the modulation format using (2.5). For the application layer system, since the mechanism of Phase II, which includes the choice of the system parameters ($L_s$, $\mu$, etc), is the same as the cross layer system, the performance difference of the two systems solely depends on the resource allocation decision.

### 4.3.2 Physical Layer Resource Allocation Algorithm

We now consider the case where only the instantaneous CSI $\tilde{H}_{k,m}[1]$ is available at the base station. To maximize the sum throughput, a conventional Multi-user diversity (MUD) algorithm will assign subcarrier $m$ to user $k^*$ with best channel gain, i.e., $k^* = \arg \max_k \left\{ |\tilde{H}_{k,m}[1]|^2 \right\}$. To ensure fairness among users, we assign subcarrier $m$ to the user $k^*$ such that

$$k^* = \arg \max_k \left\{ \frac{|\tilde{H}_{k,m}[1]|^2}{|\tilde{H}_k[1]|^2} \right\}$$

(4.19)

where $|\tilde{H}_k[1]|^2 \triangleq \frac{1}{M_c} \sum_{m=1}^{M_c} |\tilde{H}_{k,m}[1]|^2$ is the empirical average channel gain of user $k$ at the beginning of the GOP.

We introduce the following definitions that will be used in the physical layer resource allocation algorithm.

**Definitions I:**

a) Define $\Lambda$ as the set of users who are eligible for being assigned additional subcarriers.

b) Define $\Theta$ as the set of users who have not yet been assigned any subcarrier in the iteration. We design the algorithm such that each user will get at least one subcarrier.

c) Define $\Gamma$ as the set of subcarriers whose allocation decision has not yet been
made.

d) Similar to the application layer optimization algorithm, let $L_k$ be the number of subcarriers user $k$ is assigned.

For a system with coherence bandwidth larger than the subcarrier bandwidth, i.e., $\Psi > 1$, the MUD based algorithm proposed in [30] [31] allocates subcarriers in chunks, i.e., if a given user is assigned a particular subcarrier, that user will also get all the other subcarriers in the chunk. For a system using MUD with large $\Psi$, since individual users could get multiple chunks with large bandwidth, the resource allocation might be very unbalanced and the average video performance will degrade. To control the degree of imbalance in the number of subcarriers that users receive, we impose set of thresholds of $\psi_n, n = 1, 2...K - 1$, such that the sum of subcarriers for any group of $n$ users will not exceed $\psi_n$. We set $\psi_n$, for $1 \leq n \leq K - 1$, equal to

$$
\psi_n = \psi_{n-1} + \left[ \epsilon \left( \frac{M - \psi_{n-1}}{K - (n - 1)} \right) \right]
$$

(4.20)

where, for $n = 1$, this expression reduces to $\psi_1 = \left\lceil \epsilon \frac{M}{K} \right\rceil$. The parameter $\epsilon$ is chosen to be greater than or equal to 1, and controls the imbalance of the resource allocation. A larger value of $\epsilon$ means that the resource allocation decision will be more unbalanced, biased to the users who have larger channel gains. For each individual user, the number of subcarriers threshold $\psi_1$ is set to be $\epsilon$ times larger than the average number of subcarriers per user, $M_c/K$. Assuming that one user has already been assigned the maximum of $\psi_1 = \left\lceil \epsilon \frac{M_c}{K} \right\rceil$ subcarriers, the average number of subcarriers for the remaining $(K - 1)$ users is given by $(M_c - \psi_1)/(K - 1)$ and the resource for any combination of two users is limited by $\psi_1 + \left\lceil \epsilon (M_c - \psi_1)/(K - 1) \right\rceil$ subcarriers. We repeat this process iteratively for $n \leq (K - 1)$, and the total number of subcarriers assigned to any group of $n$ users can be found iteratively using (4.20). As a specific example, consider a system with 1000 subcarriers, 3 users, and $\epsilon = 1.5$. The threshold would be $\psi_1 = 500$ subcarriers for any individual user, and $\psi_2 = 875$ subcarriers for any group of two users. When the coherence bandwidth is equal to the entire bandwidth, the user with the strongest channel gain will get 500 subcarriers. The user with second best channel gain gets 375
subcarriers. The remaining 125 subcarriers are assigned to the third user. When
the coherence bandwidth becomes smaller, it will be increasingly unlikely that the
total number of subcarriers for a group of $n$ users will reach the threshold of $\psi_n$.

**Physical Layer Optimization Algorithm:**

**Step 1 Initialization:** We initialize $\Lambda$ and $\Theta$ as the complete set of users,
i.e., $\Lambda = \{1, 2, \ldots, K\}$, $\Theta = \{1, 2, \ldots, K\}$, $\Gamma$ as the complete set of subcarriers $\Gamma = \{1, 2, \ldots, M_c\}$ and $\psi_n$ is given by (4.20).

**Step 2 Subcarrier Assignment:** We choose the best channel gain from all the possible assignments,

$$(k^*, m^*) = \arg \max_{k \in \Lambda, m \in \Gamma} \left\{ \left| \frac{\tilde{H}_{k,m}[1]}{H_k[1]} \right|^2 \right\} \quad (4.21)$$

and assign subcarrier $m^*$ to user $k^*$. We update $\Gamma = \Gamma \setminus m^*$. If $k^* \in \Theta$, we update $\Theta = \Theta \setminus k^*$, meaning that user $k^*$ has been assigned at least one subcarrier. Here, $\left| \tilde{H}_{k^*,m^*}[1] \right|^2$ stands for the best channel response of all possible subcarrier assignment combinations at the current step.

**Step 3 Status Update:** We check the remaining resource and conduct the following two updates:

1) For every $n$ ($1 \leq n \leq K - 1$), we compare the sum of subcarriers for all groups of $n$ users with $\psi_n$. If the sum is equal to $\psi_n$ for any group, all the users in that group will be excluded from $\Lambda$.

2) We then check the relation between the number of subcarriers left and the cardinality of $\Theta$. To ensure that each user can get at least one subcarrier, if $|\Gamma| = |\Theta|$, we will terminate the algorithm by assigning exactly one of the remaining unallocated subcarriers to each of the users who has no subcarrier.

We then go back to Step 2 and repeat (4.21) to assign subcarriers until $\Gamma$ is empty.

4.3.3 Cross Layer Resource Allocation Algorithm

When both the application layer RD and the physical layer CSI are used for resource allocation, a cross layer algorithm attempts to satisfy the two goals of
giving more subcarriers to users with demanding RD curves and giving subcarriers to users with high channel gains. To solve the optimization problem (4.16), we propose an iterative algorithm as follows:

**Definitions II:**

a) We define $\theta_m^{(i)}$ to be the user who is assigned subcarrier $m$ at the $i$-th iteration. For example, $\theta_1^{(2)} = 3$ means user 3 is assigned subcarrier 1 at the second iteration of the algorithm. Also, similar to the physical layer optimization algorithm, let $\Lambda$ be the set of users who are eligible to be assigned additional subcarriers. We further define $|\Lambda|$ as the cardinality of the set.

b) Define $A_k^{(i)}$ to be the set of subcarriers assigned to user $k$ at the $i$-th iteration.

c) Define $\Delta_{k,m} \geq 0$ as the absolute value of the video distortion change of user $k$ by gaining or losing subcarrier $m$.

**Cross Layer Optimization Algorithm:**

**Step 1 Initialization:** We initialize the resource allocation by assigning each subcarrier to the user, $\theta_m^{(0)} = \arg \max_k \left\{ \left| \tilde{H}_{k,m}[1] \right|^2 / \left| \tilde{H}_k[1] \right|^2 \right\}$, and let $\Lambda$ be the set of all users, $\Lambda = \{1, 2...K\}$.

**Step 2 Rate and Slope Calculation:** Knowing the subcarriers in $A_k^{(i)}$, user $k$ can calculate the anticipated number of information bits to be transmitted based on the current allocation decision

$$R_k^{(i)} = \lambda \sum_{A_k^{(i)}} R_{k,m} \left( P^d, \tilde{H}_{k,m}[1] \right)$$

(4.22)

$R_k^{(i)}$ can be viewed as the aggregate rate of user $k$ across all $M_c$ subcarriers. The absolute value of the slope on the RD curve at $R_k^{(i)}$ bits comes from (4.14):

$$S_k^{(i)} = \frac{b_k}{\left( R_k^{(i)} + c_k \right)^2}$$

(4.23)

We then pick the user with the steepest slope $k^* = \arg \max_{k \in \Lambda} \{ S_k^{(i)} \}$ and consider switching one subcarrier from some other user to user $k^*$.

**Step 3 Subcarrier Reassignment:** We now check every subcarrier $m$ which is not currently assigned to user $k^*$, and consider the possibility of changing
the assignment of \( m \) from user \( \theta_m^{(i)} \) to user \( k^* \). The loss of subcarrier \( m \) will cause the MSE value of user \( \theta_m^{(i)} \) to increase by

\[
\Delta_{\theta_m^{(i)},m} = \frac{b_{\theta_m^{(i)}}}{\lambda \sum_{A_{\theta_m^{(i)},m}^{(i)}} R_{\theta_m^{(i)},m} \left( P_d, \widetilde{H}_{\theta_m^{(i)},m}[1] \right) + c_{\theta_m^{(i)}}} - \frac{b_{\theta_m^{(i)}}}{\lambda \sum_{A_{\theta_m^{(i)},m}^{(i)}} R_{\theta_m^{(i)},m} \left( P_d, \widetilde{H}_{\theta_m^{(i)},m}[1] \right) + c_{\theta_m^{(i)}}}
\]

(4.24)

and allow the MSE of user \( k^* \) to decrease by

\[
\Delta_{k^*,m} = \frac{b_{k^*}}{\lambda \sum_{A_{k^*,m}^{(i)}} R_{k^*,m} \left( P_d, \widetilde{H}_{k^*,m}[1] \right) + c_{k^*}} - \frac{b_{k^*}}{\lambda \sum_{A_{k^*,m}^{(i)}} R_{k^*,m} \left( P_d, \widetilde{H}_{k^*,m}[1] \right) + c_{k^*}}
\]

(4.25)

We then find the subcarrier \( m^* \) which maximizes the difference between \( \Delta_{k^*,m} \) and \( \Delta_{\theta_m^{(i)},m} \).

If \( \left( \Delta_{k^*,m} - \Delta_{\theta_m^{(i)},m} \right) > 0 \), we reassign subcarrier \( m^* \) to user \( k^* \) at iteration \( i + 1 \), i.e., \( \theta_m^{(i)} = k^* \), and then go back to Step 2 to update \( k^* \).

If \( \left( \Delta_{k^*,m} - \Delta_{\theta_m^{(i)},m} \right) \leq 0 \), which means that the sum of distortions will not be reduced by reassigning any subcarrier to user \( k^* \), we exclude \( k^* \) from \( \Lambda \), \( \Lambda = \Lambda \setminus k^* \) and user \( k^* \) will not be assigned any additional resource. Next, we check the cardinality of \( |\Lambda| \). If \( |\Lambda| = 1 \), meaning that there is no possibility of reassignment, we stop, otherwise we go back to Step 2 with iteration index \( i \) incremented.

The cross layer resource allocation algorithm is designed such that users with large slope are given the priority to be assigned additional subcarriers. In the algorithm, we first assign subcarriers using a MUD algorithm and estimate the number of bits that the channel can support using (4.22). From the application layer point of view, it is most likely that the sum of distortions can be reduced by assigning more subcarriers to the user with the steepest slope, so we give the priority to the user with largest slope. We test the possible reassignment of subcarriers and use the CSI to find the subcarrier which can maximize the reduction of the
sum of distortions through reassignment. After reassignment, we update the user with the steepest slope and continue the iteration until we exhaust all possibilities of subcarrier switching.

4.4 System Parameter Optimization

We study a system with 16 independently faded subcarriers and 3 users. We use a video with a resolution of 352 × 240. The video consists of 150 frames at 30 frames/second, and is organized into GOPs of 15 frames (IPPP). The content of the video includes both high motion segments and low motion segments. Each user is assigned the same video, but with random starting points. The simulation runs for one cycle of the entire video sequence, and users are then assigned another random set of starting points for the next cycle. By assigning random starting points of the same cyclic video to different users, we create instantaneous application layer diversity among users and yet have the same average complexity over time for the different users. Between cycles, different realizations of starting points will generate different levels of application diversity. We encode the video using H.264/SVC reference software JSVM version 9.19.12. For the MGS layer, the 4 × 4 DCT coefficients of each macroblock are split using MGS vector [1, 1, 2, 2, 2, 8] [54]. To specify the $a_k$, $b_k$ and $c_k$ values of RD curves in (4.14), we extract bitstreams at 20 different encoding rates from 60 to 1200 kbps offline and use these operational points to find the RD function by non-linear regression. At the decoder side, the bit stream after the first channel error is discarded. In a very crowded system, it is possible that some users will not be allocated any subcarriers and the transmission rate for the GOP is zero. The last frame of the previous GOP is held over for the duration of the current GOP.

For the physical layer, we consider a single cell of radius equal to 50 meters. The bandwidth for each subcarrier is 100kHz. The channel response consists of both path loss and multipath fading, and the amplitude squared of the multipli-
cative channel coefficient $H_{k,m}[l]$ is given by

$$|H_{k,m}[l]|^2 = |\alpha_{k,m}[l]|^2 \cdot K_0 \cdot \left(\frac{d_0}{d_k}\right)^\beta$$

(4.26)

where $\alpha_{k,m}[l]$ follows the Jakes’ model and is generated using the statistical model proposed by [62]. Also, $d_k$ is the distance of user $k$ to the base station and $d_0 = 10\text{m}$ is the reference distance. The path-loss model is accurate when $d_k > d_0$ [4]. The users are perfectly power controlled and the average received power is $17.8\text{ dB}$ per subcarrier. We set the path-loss exponent $\beta = 3$ and $K_0 = -30\text{dB}$ is a constant.

For all three optimization schemes, we apply a rate $u = 1/2$ convolutional code with code generator polynomial [23, 35] in octal. The codeword length is equal to the length of the entire GOP bitstream, and the coded bits are interleaved across different subcarriers. The value of $SER_t$ is set to 0.15. At the receiver, we use soft-decision decoding with eight reliability ranges. The MQAM modulation format can be $\{M = 4, 8, 16, 32, 64, 128, 256\}$ for all three algorithms. When the channel is in deep fade such that $R_{k,m}(P_{k,m}^d, \bar{H}_{k,m}[iL_s + 1]) < 2$ in (4.8), the user will choose to transmit nothing for the duration of the time epoch. We also ignore the effect of inter-carrier interference in the simulation. For the physical layer optimization algorithm, we set $\epsilon = 1.5$, so one user cannot be assigned more than 150% of the average number of subcarriers. Because of the limitation of the computation power, for a specific normalized Doppler spread, the channel is generated independently GOP by GOP. With the GOP length equals to $5 \times 10^4$ thousands symbols, the simulation is accurate for $f_{nd} > 10^{-4}$, and as the normalized Doppler spread decreases, our numerical results become progressively less accurate.

### 4.4.1 System With Different $L_s$

In Fig. 4.2, we show the performance of the three resource allocation algorithms with respect to different Doppler spreads. The pilot insertion space $L_s$ equals 100 and the number of pilots used for channel estimation $K_e$ is 24. The ratio $\mu$ between pilot and the power equals unity. The solid lines are the numerical results obtained from the RD curves and can be considered as error-free distortion,
or distortion at the encoder side. The calculation of the PSNR values for these curves consists of the distortion caused by source compression only. The dashed lines are the distortion results measured at the decoder; the distortion consists of both the degradation caused by source compression and that caused by channel errors. The effects of packet loss, errors in RD curve fitting, and imperfection of encoder rate control are included in the simulation.

Figure 4.2: \( L_s = 100, 16 \) Subcarriers, 3 Users, \( SER_t = 0.15, \mu = 1 \). The PSNR performance for a system with \( f_{nd} > 10^{-2} \) is less than 15dB.

For all three algorithms, the gap between the curves at the encoder side and decoder side is small when the normalized Doppler spread is between \( 10^{-8} \) and \( 10^{-4} \). In Fig. 4.3, we show the decoded BER conditioning on the user is transmitting information bits, i.e., \( R_{k,m}(P_{k,m}^d, \tilde{H}_{k,m}[iL_s + 1]) \geq 2 \) in (4.8). When \( 1/f_{nd} \) is significantly larger than \( L_s = 100 \), the channel is relatively constant within each duration of \( L_s \) symbols. The decoded bit error rate for this scenario is in the range of \( 10^{-7} \) to \( 10^{-6} \). Since we update the modulation format every \( L_s \) symbols, the channel stays relatively constant within the duration of the \( L_s \) symbols and the
channel estimation error is small. The actual raw SER every group of 100 symbols is similar to the SER of an AWGN channel and is close to the SER estimated using (2.4). In addition, since the channel is varying slowly, the correlation between the data symbols and pilot symbols is relatively high, and the channel estimation error given in (4.9) is small. In other words, for $L_s = 100$ and $f_{nd} < 10^{-4}$, the proposed mechanism of PSAM (Phase II) can accurately adapt to the variation of the channel and properly control the channel error rate.

![Decoded BER vs $f_{nd}$, $L_s = 100$, $SER_t = 0.15$](image)

**Figure 4.3:** Decoded BER vs $f_{nd}$, $L_s = 100$, $SER_t = 0.15$

When $f_{nd} = 10^{-3}$, we see that the curves at the encoder side and decoder side start to diverge. Since the channel estimates becomes increasingly outdated, the modulation format chosen based upon the CSI becomes increasingly meaningless. We see a significant increase of the decoder BER when $f_{nd}$ reaches $10^{-3}$, and the gap between the curves at the encoder side and decoder side widens. When the normalized Doppler spread is $10^{-2}$, the decoded BER is too large for the system to function.

In Fig. 4.4, we decrease the value of $L_s$ to 25. Recall that the fraction of data is $\lambda = (L_s - 1)/L_s$. Decreasing the value of $L_s$ will increase the amount of overhead (pilot symbols) used for channel estimation. We thus observe a decrease
in the performance for all three algorithms at the encoder side as the number of source bits transmitted decreases. However, decreasing the value of \( L_s \) will help the system to achieve higher estimation accuracy at high Doppler, and the gap between the performance at the encoder side and the decoder side is narrower compared to Fig. 4.2 at \( f_{nd} = 10^{-3} \). Despite the drop of the source rate, we see that for a system with \( f_{nd} = 10^{-3} \), a pilot spacing \( L_s = 25 \) has better performance than when \( L_s = 100 \). If we further decrease the value of \( L_s \) to 5 as in Fig. 4.5, the loss of the source data further increases to 20\%, but the benefits of accurate channel estimation and modulation adaptation allow the system to operate at a reasonable PSNR value at \( f_{nd} = 10^{-2} \) at the decoder side.

### 4.4.2 Systems with Different Resources

We now study the performance for different parameter values. In Fig. 4.6, we increase the number of users to 4. In Fig. 4.7 and Fig. 4.8, we show the performance of a system of 24 subcarriers with 3 and 4 users respectively. Comparing Figures 4.2, 4.6, 4.7 and 4.8, we see that the system with 24 subcarriers and 3
Figure 4.5: $L_s = 5$, 16 Subcarriers, 3 Users, $SER_t = 0.15$, $\mu = 1$

users not only has the best performance, but also has the largest gap between the cross layer and physical layer optimization algorithms. This is because when the system has 24 subcarriers and 3 users, the cross layer algorithm has more degrees of freedom to allocate the resources among the users.

In Table 4.1, we show a typical example of the performance evolution of a three-user system with 16 subcarriers and $f_{nd} = 10^{-6}$. The resource allocation is done based on the estimate of the first sample of the GOP, and we show the change of the estimated MSE for each individual user at each step of the iteration. Among the three users, the first and second users have demanding RD curves and the corresponding $b_k$ of the RD information is much larger than that of the third user. For the cross layer algorithm initialization, the average PSNR is equal to 32.48 dB, and subcarriers are assigned evenly to all the users. The algorithm will converge in 3 steps (3 iterations) and the final subcarrier allocation is 6 and 7 for the two demanding users and 3 for the third user. The PSNR value of the cross layer algorithm is 33.16 dB. In this case, the cross layer optimization improves the estimated performance of the system by only 0.7 dB.
In Table 4.2, we study a system with 24 subcarriers and 3 users. To compare with the previous example fairly, the channel realizations for 16 out of 24
subcarriers are the same as the realizations of the previous example. Realizations for the remaining 8 subcarriers are generated using the same mechanism as for the first 16 subcarriers. For a 24-subcarrier system, the initialization of the allocation is (7,8,9) subcarriers to the three users, and the corresponding average PSNR is 34.82 dB. It takes the cross layer algorithm 5 switches of subcarriers, and the cross layer algorithm assigns (9,10,5) subcarriers to the three users. The resulting average PSNR performance is 36.17 dB, and so the cross layer algorithm improves the performance by 1.35 dB. Going from 16 to 24 subcarriers, we see that it takes the

Table 4.1: Evolution of the performance: 16 Subcarriers/3 Users

<table>
<thead>
<tr>
<th></th>
<th>U1 MSE</th>
<th>U1 NSub</th>
<th>U2 MSE</th>
<th>U2 NSub</th>
<th>U3 MSE</th>
<th>U3 NSub</th>
<th>Avg. MSE</th>
<th>PSNR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initialization</td>
<td>53.68</td>
<td>5</td>
<td>56.52</td>
<td>5</td>
<td>~0</td>
<td>6</td>
<td>36.73</td>
<td>32.48</td>
</tr>
<tr>
<td>Step 1</td>
<td>53.68</td>
<td>5</td>
<td>45.89</td>
<td>6</td>
<td>~0</td>
<td>5</td>
<td>33.19</td>
<td>32.97</td>
</tr>
<tr>
<td>Step 2</td>
<td>44.82</td>
<td>6</td>
<td>45.89</td>
<td>6</td>
<td>4.66</td>
<td>4</td>
<td>31.79</td>
<td>33.10</td>
</tr>
<tr>
<td>Step 3</td>
<td>44.82</td>
<td>6</td>
<td>40.35</td>
<td>7</td>
<td>9.08</td>
<td>3</td>
<td>31.41</td>
<td>33.16</td>
</tr>
</tbody>
</table>
Table 4.2: Evolution of the performance: 24 Subcarriers/3 Users

<table>
<thead>
<tr>
<th></th>
<th>U1 MSE</th>
<th>U1 NSub</th>
<th>U2 MSE</th>
<th>U2 NSub</th>
<th>U3 MSE</th>
<th>U3 NSub</th>
<th>Avg. MSE</th>
<th>PSNR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initialization</td>
<td>34.38</td>
<td>7</td>
<td>29.94</td>
<td>8</td>
<td>~ 0</td>
<td>9</td>
<td>21.43</td>
<td>34.82</td>
</tr>
<tr>
<td>Step 1</td>
<td>29.41</td>
<td>8</td>
<td>29.94</td>
<td>8</td>
<td>~ 0</td>
<td>8</td>
<td>19.78</td>
<td>35.17</td>
</tr>
<tr>
<td>Step 2</td>
<td>25.27</td>
<td>9</td>
<td>29.94</td>
<td>8</td>
<td>~ 0</td>
<td>7</td>
<td>18.40</td>
<td>35.48</td>
</tr>
<tr>
<td>Step 3</td>
<td>25.27</td>
<td>9</td>
<td>25.31</td>
<td>9</td>
<td>~ 0</td>
<td>6</td>
<td>16.86</td>
<td>35.86</td>
</tr>
<tr>
<td>Step 4</td>
<td>23.11</td>
<td>10</td>
<td>25.31</td>
<td>9</td>
<td>~ 0</td>
<td>5</td>
<td>16.14</td>
<td>36.05</td>
</tr>
<tr>
<td>Step 5</td>
<td>23.80</td>
<td>9</td>
<td>23.31</td>
<td>10</td>
<td>~ 0</td>
<td>5</td>
<td>15.71</td>
<td>36.17</td>
</tr>
</tbody>
</table>

cross layer algorithm more steps to converge and the PSNR improvement is larger. Compared with a 16-subcarrier system, a 24-subcarrier system allows the cross layer optimization to more flexibly perform subcarrier switching and thus have a larger gain.

4.4.3 Systems With Different $\mu$

Besides lowering the value of $L_s$, another way of achieving accurate channel estimation is to allocate more power to pilot symbols. In Fig. 4.9, we show the system performance versus the pilot-to-data power ratio $\mu$ for the cross layer algorithm for $f_{nd} = 10^{-3}$. When we increase $\mu$, when $L_s$ is 100, the power of the data symbols is not significantly smaller than for the system with $\mu = 1$, and we thus do not observe a large degradation of performance at the encoder side. On the other hand, the channel estimation is more accurate with large channel pilot power, and the decoder PSNR improves.

In Fig. 4.10, we decrease $L_s$ to 25. Compared to the previous plot, since the fraction of pilots increases, if we increase their power, the impact to the data power will be larger. At the encoder, we see a 0.25 dB performance drop if $\mu$ increases from 0.5 to 6. As $\mu$ increases, the average power of the data decreases and users will statistically choose more conservative modulation formats. At the decoder, when pilot power is small, the channel estimation accuracy is bad and the performance...
gap between encoder and decoder is more than 2.5 dB when $\mu = 0.5$. This gap narrows if we increase $\mu$. When $\mu$ is larger than 2, the choice of modulation format at the encoder becomes the dominant effect and the performance at the decoder starts to drop. The best performance at the decoder side is achieved when $\mu$ is in the range from 1 to 2 in this particular setting.

![Figure 4.9: $f_{nd} = 10^{-3}$, $L_s = 100$, 16 Subcarriers, $SER_t = 0.15$](image)

### 4.5 System Capacity Analysis

We now study the system performance when the number of users increases. We fix the number of subcarriers to 16 and change the number of users in the system from 3 to 9. In Fig. 4.11, we show the average performance of all users when $f_{nd} = 10^{-6}$ and $L_s = 100$. We see that when $f_{nd}$ is small, the gain of using the cross layer scheme is very large compared to the application layer scheme. With 9 users, the users will normally get a very small number of subcarriers. For the
application layer algorithm, the subcarriers are randomly assigned, and because the number of subcarriers for each user is small or zero, it is very likely that at least one user will be in a very bad situation. The user will thus get zero rate with high probability, and this has a large impact on the average distortion of the group.

Compared to the physical layer algorithm, the PSNR gain of the cross layer scheme is 1dB in a three-user system and more than 2 dB in a 6 or 9-user system. For a fixed average PSNR of 30 dB, the capacity gain for the cross layer algorithm is 3 times compared to the application layer algorithm and about 2 times compared to the physical layer algorithm.

In Fig. 4.12, we let $f_{nd} = 10^{-3}$ and $L_s = 25$. We see that when $f_{nd}$ is large, the cross layer scheme outperforms the physical layer and application layer schemes by about 1 dB and 0.5 dB, respectively, with 3 users. The gap between the cross layer and the physical layer algorithms is always about 1 dB when the
$f_{nd}$ increases, which is smaller than that for $f_{nd} = 10^{-6}$. This is because when $f_{nd}$ increases, the CSI used for resource allocation is outdated more quickly, and this affects the ability of the cross layer algorithm to balance the resources among the users. In other words, at high $f_{nd}$, the overall system performance is more determined by the number of subcarriers assigned to each user, and the effect of the CSI of the first symbol of the GOP has less impact than that for a slow fading system. On the other hand, the restrictions of (4.20) will not allow the resource allocation decision to be significantly unequal across the users. The cross layer optimization algorithm cannot perform many subcarrier switchings in a crowded system, and the gap between the cross layer and physical layer algorithms is thus smaller.

Comparing the cross layer and application layer optimization algorithms, we see that the performance of the application layer algorithm decreases sharply when the number of users increases, because some of the less demanding users in the system will not be allocated any subcarriers. The performance of these users
Figure 4.12: $L_s=25$, $f_{nd} = 10^{-3}$, 16 Subcarriers, $SER_t = 0.15$, $\mu = 1$

will be measured by holding the last frame of the previous GOP, resulting in a low PSNR value.

4.6 Conclusion

In this chapter, we extended the resource allocation problem of Chapter 3 to a system with arbitrary normalized Doppler spread. We use both the application layer RD information and the physical layer CSI to allocate subcarriers to video users. After the resource allocator assigns the subcarriers to the users, the users periodically send a pilot symbol and update the modulation format of each subcarrier. The repetition interval for pilot symbols, $L_s$, is a critical parameter to control the tradeoff between the channel outdating and source rate. We illustrated the sensitivity of the system to different values of $L_s$. When the Doppler spread is smaller, the physical layer algorithm can provide high throughput to the system and thus outperforms the application layer algorithm. When the Doppler
spread is large, one should rely more on the application layer information (RD function) for resource allocation as the instantaneous CSI will be outdated soon after the allocation decision and the throughput of each subcarrier will tend to be the same. While the physical layer algorithm is sometimes better than the application layer algorithm, and the application layer algorithm is sometimes better, the cross layer algorithm has robust performance across different Doppler spreads because it always outperforms the two baseline algorithms.

Chapter 4, in part is currently being prepared for submission for publication by Dawei Wang, Laura Toni, Pamela Cosman and Laurence Milstein "Resource Allocation and Performance Analysis for Multiuser Video Transmission over Doubly Selective Channels", IEEE Transaction on Wireless Communications. The dissertation author is the primary investigator and author of this paper.
Chapter 5

Conclusions and Future Work

In this dissertation, we investigated the problem of resource allocation for multiple video users in an uplink multi-carrier setting. In a bandwidth and power constrained system, multiple video users with different rate distortion functions would like to send compressed video data to the base station. We designed bandwidth sharing schemes such that both the source characteristics and the channel quality are considered to improve the resource allocation decisions.

In Chapter 3, we assumed a slow fading environment with the normalized Doppler spread $f_{nd} = 0$. As the channel realization stays constant for the duration of a GOP, the throughput of each subcarrier can be estimated accurately before transmission. We proved that in a continuous channel setting, the optimal resource allocation should assign frequency bands to users based on a criterion which considers the product of the slope of the rate distortion function and a physical layer metric related to water-filling. Based on this concept, we designed a cross layer resource allocation algorithm using a block fading model. The simulation results showed that the cross layer scheme can robustly improve the average performance of the system compared to resource allocation algorithms using only a single layer of information.

In Chapter 4, we generalized our resource allocation to systems with arbitrary mobility. We incorporated a pilot symbol assisted modulation scheme to combat the channel estimation error and channel outdating for a fast fading chan-
nel. We showed that the performance of the system at different Doppler spreads can be controlled by the frequency of pilot insertion. We then characterized the tradeoff between the channel outdating and pilot insertion spacing. We also showed a crossover of the performance of two baseline algorithms using one layer of information for resource allocation when the Doppler changes. When the Doppler spread is small, the physical layer algorithm can provide high throughput to the system and thus outperform the application layer algorithm. When the Doppler spread is large, one should rely more on the application layer information for resource allocation as the instantaneous CSI will be outdated soon after the allocation decision and the throughput of each subcarrier will tend to be the same. We finally compared the system capacity of the three different algorithms at different degrees of user mobility.

Possible future work includes:

- In Chapter 4, we used one resource allocation decision for each GOP and then updated the modulation format using PSAM. One possible extension can be multiple resource allocations within the duration of a GOP. Since the allocation algorithm can be computationally intensive, the resource allocation might not be able to be updated every coherence time. However, frequent resource allocation can reduce the effect of channel outdating and potentially better exploit the multiuser channel diversity.

- In this dissertation, the resource allocation is done on a GOP-to-GOP basis. In a delay insensitive scenario, if the users are allowed to buffer multiple GOP’s, the system is then able to better exploit the diversity of different fades of the time varying channel at the resource allocation level.

- We focused only on multiuser resource allocation in this dissertation. A joint optimization with both multiuser resource allocation and single user transmission optimization (e.g., UEP) needs to be studied.
Appendix A

Performance Change Calculation
For Reassigning Subcarriers

Consider a user $k$ who gets assigned a set of $A^{(i)}_k$ subcarriers. As discussed in Step 2 of Section 3.3, the optimal power allocation scheme is

$$P^*_k,m = \left[ \frac{1}{\lambda_k} - \frac{1}{\eta |H_{k,m}|^2} \right]^+ \quad (A.0.1)$$

$\forall m \in A^{(i)}_k$. We want to find the video performance degradation of user $k$ after losing a subcarrier $\hat{m}$, $\hat{m} \in A^{(i)}_k$, $\frac{1}{\lambda_k} > \frac{1}{\eta |H_{k,\hat{m}}|^2}$. For the scenario that all the subcarriers’ frequency responses are below the water level, or $\frac{1}{\lambda_k} > \frac{1}{\eta |H_{k,m}|^2}$, the operating rate (in bits/symbol) of user $k$ is given by

$$r^*_k = \sum_{m \in A^{(i)}_k} \log_2 \left[ 1 + \eta |H_{k,m}|^2 \left( \frac{1}{\lambda_k} - \frac{1}{\eta |H_{k,m}|^2} \right) \right] \quad (A.0.2)$$

Note that we start the resource allocation by assigning the subcarrier to the user with the best response, so we expect that the condition of $\frac{1}{\lambda_k} > \frac{1}{\eta |H_{k,m}|^2}$ holds for most of the subcarriers at the beginning of the iterations. The video distortion is

$$D_k = a_k + \frac{b_k}{\sum_{m \in A^{(i)}_k} \log_2 \left[ 1 + \eta |H_{k,m}|^2 \left( \frac{1}{\lambda_k} - \frac{1}{\eta |H_{k,m}|^2} \right) \right] + c_k} \quad (A.0.3)$$
After losing subcarrier \( \hat{m} \), the water level will increase by

\[
\frac{\left( \frac{1}{\lambda_k} - \frac{1}{\eta |H_k, \hat{m}|^2} \right)}{\left( |A_k^{(i)}| - 1 \right)}
\]  

(A.0.4)

and the updated video distortion is expressed as

\[
\tilde{D}_k = a_k + \frac{b_k}{\sum_{m \in \left( A_k^{(i)} - \hat{m} \right)} \log_2 \left[ 1 + \eta |H_{k,m}|^2 \left( \frac{|A_k^{(i)}|}{|A_k^{(i)}| - 1} \right)^{\lambda_k} - \frac{1}{\eta |H_{k,m}|^2} - \frac{1}{\eta |H_{k,m}|^2 \left( |A_k^{(i)}| - 1 \right)} \right]} + c_k
\]

(A.0.5)

We can then calculate the performance change of user \( k \) for losing subcarrier \( \hat{m} \) as \(-\Delta_{k,\hat{m}} = \tilde{D}_k - D_k\). If user \( k \) is given one subcarrier, the performance improvement for that user can be found in a similar way.
Bibliography


