

## Efficace utilisation des ressources de CQI Feedback pour les systèmes sans fil multi-utilisateur multi-porteuse

Mohammad Abdul Awal

#### ▶ To cite this version:

Mohammad Abdul Awal. Efficace utilisation des ressources de CQI Feedback pour les systèmes sans fil multi-utilisateur multi-porteuse. Computers and Society [cs.CY]. Université Paris Sud - Paris XI, 2011. English. NNT: 2011PA112223 . tel-00636659

### HAL Id: tel-00636659 https://theses.hal.science/tel-00636659

Submitted on 28 Oct 2011

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers. L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

#### Efficient CQI Feedback Resource Utilization for Multi-user Multi-Carrier Wireless Systems

by

Mohammad Abdul Awal

A dissertation submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

in

Computer Science

in the

#### GRADUATE DIVISION

#### of the

#### UNIVERSITY OF PARIS-SUD XI

#### Committee in charge:

Advisor:

Lila Boukhatem Associate Professor, University of Paris-Sud XI, France

#### **Reviewers:**

André-Luc Beylot Thierry Turletti

eylot Professor, IRIT - ENSEEIHT, France cletti Research Director, INRIA, Sophia Antipolis, France

#### Members of the jury:

Khaldoun Al Agha Tijani Chahed Salah Eddine Elayoubi

Professor, University of Paris-Sud XI, France Professor, Telecom SudParis, France Senior Researcher, Orange Labs, France

 $26 \ {\rm October} \ 2011$ 

To my parents, wife, and son.

ii

#### Résumé

La technologie OFDMA (Orthogonal frequency division multiple access) a été adoptée par les systèmes de télécommunications de 4ème génération (4G) comme technique de transmission et d'accès multiple pour ses performances supérieures en termes d'efficacité spectrale. Dans ce type de systèmes, l'adaptation dynamique du débit en fonction de la qualité du canal CQI (Channel Quality Indicator) constitue une problématique de recherche d'actualité qui attire l'attention de plusieurs acteurs académiques et industriels. Ce problème d'adaptation dynamique est encore plus complexe à gérer dans des environnements multi-utilisateurs hétérogènes et à ressources limitées tels que les systèmes OFDMA comme WiMAX Mobile et Longterm Evolution (LTE). Dans cette thèse, nous nous intéressons au problème d'allocation de ressources de l'information de feedback relative au CQI dans le cadre de systèmes OFDMA multi-porteuses multi-utilisateurs. Dans le but de réduire la charge (overhead) du feedback, nous proposons une méthode de prédiction du CQI basée sur l'exploitation de la corrélation temporelle de ce dernier et d'une solution inter-couches. L'objectif est de trouver des schémas d'allocation de ressources adaptatifs respectant les contraintes de qualité de service (QoS) applicatives.

Nous proposons en premier lieu un algorithme de réduction de feedback PBF (Prediction Based Feedack) qui permet à la station de base (BS) à prédire certaines occurrences du CQI en se basant sur l'algorithme des moindres carrés récursif RLS (Recursive least-square). Les résultats de simulation montrent que l'outil de prédiction du CQI réduit sensiblement l'overhead du feedback et améliore par conséquent le débit de la liaison montante. Nous proposons, par la suite, une version opportuniste de PBF pour atténuer les éventuels effets de sur et sous estimations liées à l'algorithme de prédiction. Dans ce mécanisme, nous exploitons les informations inter-couches pour améliorer les performances des mécanismes de feedbacks périodiques dont PBF fait partie. L'approche opportuniste améliore sensiblement les performances du système pour les cas de mobilité élevée comparés aux cas de faible mobilité.

Dans un second temps, nous proposons une plateforme (FEREP : feedback resource allocation and prediction) basée sur une approche inter-couches. Implémentée au niveau de la station BS, FEREP intègre les fonctionnalités de prédiction, d'adaptation dynamique du CQI et d'ordonnancement des demandes de feedback. Elle comporte trois modules. Le module FWA (feedback window adaptation) gère dynamiquement la fenêtre de feedbacks de chaque station mobile (MS) en se basant sur les messages ARQ (Auto.. ) reçus qui reflètent l'état actuel des canaux respectifs. Le module PBFS (priority-based feedback scheduling) effectue ensuite l'ordonnancement des feedbacks en tenant compte de la taille de la fenêtre de feedback, du profil de l'utilisateur sous la contrainte de la limitation des ressources globales du systèmes réservées au feedback. Afin de choisir les paramètres de transmission MCS (modulation and coding schemes), le module PBF (prediction based feedback) est utilisé pour les utilisateurs dont le feedabck n'a pas pu être ordonnancé dans la trame courante. Les résultats de simulation ont montré un gain significatif des performances de FREREP en comparaison à un mécanisme de référence, en particulier, sous de fortes contraintes de limitation des ressources du feedback.

Le protocole ARQ génère un accusé de réception uniquement si l'utilisateur est sélectionné par l'ordonnanceur pour envoyer des données sur la liaison descendante. Dans le cas où la fréquence d'ordonnancement des utilisateurs sur le lien descendant est réduite, les messages ARQ s'en trouvent également réduits, dégradant par conséquent les performances de la plateforme FEREP proposée ci-dessus. En effet, dans ce cas la signalisation ARQ devient insuffisante pour adapter efficacement la fenêtre de feedback de chaque utilisateur. Pour pallier à ce problème, nous proposons l'algorithme DCRA (dynamic CQI resource allocation) qui utilise deux modes d'estimation de la fenêtre de feedback. Le premier est un mode hors-ligne basé sur des études empiriques permettant d'estimer la fenêtre moyenne optimale de feedback en utilisant les profils applicatif et de mobilité de l'utilisateur. Notre analyse de performance par simulation montre que la fenêtre de feedback peut être estimée en fonction de la classe de service des utilisateurs et de leurs profils de mobilité pour un environnement cellulaire donné. Le second mode de fonctionnement de DCRA effectue une adaptation dynamique de la fenêtre en temps réel dans le cas où la signalisation ARQ est suffisante. Une étude comparative avec les mécanismes DFS (deterministic feedback scheduling) et OFS (opportunistic feedback scheduling), a montré que DCRA arrive à réaliser un meilleur gain en ressources montantes grâce à la réduction de l'overhead des feedbacks, sans pour autant trop dégrader le débit descendant des utilisateurs. Du point de vue des utilisateurs, DCRA améliore les contraintes de QoS tels que le taux de perte de paquets et réduit la consommation énergétique des terminaux grâce à la réduction de feedback.

**Mots-clés :** OFDMA, système multiutilisateurs multiporteuses, la réduction de la feedback, l'allocation des ressources des feedbacks, feedback limitée, la prédiction des feedback CQI, le conception inter-couche, feedback du canal adaptatif, la différenciation des services.

#### Abstract

Orthogonal frequency division multiple access (OFDMA) technology has been adopted by 4th generation (a.k.a. 4G) telecommunication systems to achieve high system spectral efficiency. A crucial research issue is how to design adaptive channel quality indicator (CQI) feedback mechanisms so that the base station can use adaptive modulation and coding (AMC) techniques to adjust its data rate based on the channel condition. This problem is even more challenging in resource-limited and heterogeneous multiuser environments such as Mobile WiMAX, Long-term Evolution (LTE) networks. In this thesis, we consider CQI feedback resource allocation issue for multiuser multicarrier OFDMA systems. We exploit time-domain correlation for CQI prediction and cross-layer information to reduce feedback overhead for OFDMA systems. Our aim is find resource allocation schemes respecting the users QoS constraints.

Our study begins with proposing prediction based feedback (PBF) which allows the base station to predict the CQI feedbacks based on recursive least-square (RLS) algorithm. We showed that it is useful to use channel prediction as a tool to reduce the feedback overhead and improve the uplink throughput. Then, we propose an opportunistic periodic feedback mechanism to mitigate the possible under and over estimation effects of CQI prediction. In this mechanism, we exploited the cross-layer information to enhance the performance of periodic feedback mechanisms. The opportunistic mechanism improves the system performance for high mobility cases compared to low mobility cases.

For OFDMA systems with limited feedback resource, we propose an integrated cross-layer framework of feedback resource allocation and prediction (FEREP). The proposed framework, implemented at the BS side, is composed of three modules. The feedback window adaptation (FWA) module dynamically tunes the feedback window size for each mobile station based on the received ARQ (Automatic Repeat Request) messages that reflect the current channel condition. The priority-based feedback scheduling (PBFS) module then performs feedback allocation by taking into account the feedback window size, the user profile and the total system feedback budget. To choose adapted modulation and coding schemes (MCS), the prediction based feedback (PBF) module performs channel prediction by using recursive least square (RLS) algorithm for the user whose channel feedback has not been granted for schedule in current frame. Through extensive simulations, the proposed framework shows significant performance gain especially under stringent feedback budget constraint. ARQ protocol receives users acknowledgement only if the user is scheduled in the downlink. The reduction in users scheduling frequency also reduces the rate of ARQ hints and degrades the performance of above contributions. In this case, it is difficult to exploit the ARQ signal to adapt the feedback window for that user. To address this issue, we propose a cross-layer dynamic CQI resource allocation (DCRA) algorithm for multiuser multicarrier OFDMA systems. DCRA uses two modes for feedback window estimation. The first one is an off-line mode based on empirical studies to derive optimal average feedback window based on user application and mobility profile. Our experimental analysis shows that the feedback window can be averaged according to users service class and their mobility profile for a given cell environment. DCRA performs a realtime dynamic window adaptation if sufficient cross-layer hints are available from ARQ signaling. DCRA increases uplink resource by reducing feedback overhead without degrading downlink throughout significantly compared to deterministic feedback scheduling (DFS) and opportunistic feedback scheduling (OFS). From the users perspective, DCRA improves QoS constraints like packet loss rate and saves users power due to feedback reduction.

**Keywords:** OFDMA, multiuser multicarrier system, feedback reduction, feedback resource allocation, limited feedback, CQI feedback prediction, cross-layer design, adaptive channel feedback, service differentiation.

## Contents

| Li  | List of Figures x  |             |                                       |    |  |  |
|---|--------------------|-------------|---------------------------------------|----|--|--|
| $\mathbf{Li}$   | List of Tables xii |             |                                       |    |  |  |
| 1   | Intr               | roduction 1 |                                       |    |  |  |
|   | 1.1                | Motiva      | ation and Goals                       | 2  |  |  |
|   | 1.2                | Contri      | butions                               | 3  |  |  |
|   | 1.3                | Outlin      | e of the Thesis                       | 5  |  |  |
| <b>2</b>  | Lite               | rature      | Review                                | 7  |  |  |
|   | 2.1                | OFDM        | IA Basics                             | 7  |  |  |
|   | 2.2                | Chann       | el Quality Indicator (CQI)            | 9  |  |  |
|   | 2.3                | CQI F       | eedback Overhead                      | 10 |  |  |
|   | 2.4                | CQI F       | eedback Reduction                     | 11 |  |  |
|   |                    | 2.4.1       | Quantization                          | 11 |  |  |
|   |                    | 2.4.2       | Opportunistic Feedback                | 12 |  |  |
|   |                    | 2.4.3       | Best-M Feedback                       | 13 |  |  |
|   |                    | 2.4.4       | Channel Prediction                    | 13 |  |  |
| 3 Prediction Based CQI Feedback To Reduce Feedback Overhead |                    |             | 16                                    |    |  |  |
|   | 3.1 System Model   |             | n Model                               | 17 |  |  |
|   |                    | 3.1.1       | CQI Measurements                      | 18 |  |  |
|   |                    | 3.1.2       | Adaptive Modulation and Coding Scheme | 18 |  |  |
|   |                    | 3.1.3       | Uncertainty Period                    | 19 |  |  |
|   |                    | 3.1.4       | Automatic Repeat Request              | 20 |  |  |
|   |                    | 3.1.5       | Doppler Frequency                     | 20 |  |  |

|                                      | 3.2       | Interval Based Feedback (IBF)         |  | 20         |
|--------------------------------------|-----------|---------------------------------------|--|------------|
|                                      | 3.3       | B Prediction Based Feedback (PBF)     |  |            |
|                                      |           | 3.3.1                                 | Prediction Model   | 21         |
|                                      |           | 3.3.2                                 | Effect of Feedback Prediction Error                              | 25         |
|                                      | 3.4       | Oppor                                 | tunistic Periodic Feedback                                       | 31         |
|                                      |           | 3.4.1                                 | Opportunistic Interval Based Feedback (OIBF)                     | 32         |
|                                      |           | 3.4.2                                 | Opportunistic Prediction Based Feedback (OPBF)                   | 32         |
|                                      | 3.5       | Simula                                | tion Results   | 33         |
|                                      |           | 3.5.1                                 | Simulation Settings  | 33         |
|                                      |           | 3.5.2                                 | Performance Evaluation   | 36         |
|                                      |           | 3.5.3                                 | Space and Memory Complexity for Prediction Algorithm             | 38         |
|                                      | 3.6       | Conclu                                | sion   | 39         |
| 4                                    | ۸de       |                                       | Foodbook Decourse Allocation and Dradiction                      | 41         |
| 4                                    |           | Introdu                               | reedback Resource Anocation and Frediction                       | <b>41</b>  |
|                                      | 4.1       | Polato                                |  | 41         |
| <ul> <li>4.2 Related Works</li></ul> |           |                                       | 42   |            |
|                                      |           | and Production and Production (FFPFP) | 44   |            |
|                                      | 4.4       |                                       | Exchange Window Adoptation (FEWA)                                | 40         |
|                                      |           | 4.4.1                                 | Priority based Faedback Scheduling (DEFS) under Tatel Budget Con | 40         |
|                                      |           | 4.4.2                                 | straint  | 47         |
|                                      |           | 4.4.3                                 | Prediction Based Feedback (PBF)                                  | 48         |
|                                      | 4.5       | Perform                               | mance Metric for Feedback Window Adaptation                      | 48         |
|                                      |           | 4.5.1                                 | Interval Based Feedback (IBF) Case                               | 49         |
|                                      |           | 4.5.2                                 | Prediction Based Feedback (PBF) Case                             | 52         |
|                                      | 4.6       | Simula                                | tion Study   | 53         |
|                                      |           | 4.6.1                                 | Simulation Setting   | 54         |
|                                      |           | 4.6.2                                 | Homogenous Scenario  | 55         |
|                                      |           | 4.6.3                                 | Heterogeneous Scenario   | 57         |
|                                      | 4.7       | Conclu                                | usion  | 59         |
| ĸ                                    | D         | omia (                                | OI Resource Allocation for Oos Constrained OFDMA Sustained       | <b>В</b> 1 |
| J                                    | 5 1       | Introd                                | uction   | 61         |
|                                      | ป.⊥<br>ธา | Suctor                                |  | 60         |
|                                      | 0.2       | system                                | 1 MOUGI  | 02         |

| Bi                                  | Bibliography 83 |  |   |    |  |
|-------------------------------------|-----------------|--|---|----|--|
|                                     | 6.2             | Issues   | for Further Work  | 81 |  |
|                                     | 6.1             | Conclu   | usions  | 78 |  |
| 6 Conclusions                       |                 | ns   | 78  |    |  |
|                                     | 5.6             | Conclu   | usion   | 77 |  |
|                                     |                 | 5.5.3  | Performance of DCRA   | 73 |  |
|                                     |                 | 5.5.2  | Empirical Study: Determining Matrix $W$                                       | 72 |  |
|                                     |                 | 5.5.1  | Simulation Environment  | 71 |  |
| 5.5 Simulation Results and Analysis |                 | ation Results and Analysis                             | 70  |    |  |
|                                     |                 | 5.4.2  | EMw-DCRA: DCRA with Empirical Feedback Window $\hfill \ldots \ldots \ldots$ . | 69 |  |
|                                     |                 | 5.4.1  | RTw-DCRA: DCRA with Realtime Feedback Window Adaptation $\ . \ .$ .           | 66 |  |
|                                     | 5.4             | Dynar  | nic CQI Resource Allocation (DCRA)  | 65 |  |
|                                     |                 | 5.3.2  | CQI Feedback Scheduling Strategy  | 64 |  |
|                                     |                 | 5.3.1  | Downlink Proportional Fair Scheduling (PFS)                                   | 63 |  |
|                                     | 5.3             | Downlink Scheduling and Feedback Allocation Strategy 6 |   |    |  |

# List of Figures

| 2.1  | Frame structure of an OFDMA system (IEEE 802.16e)  | 8  |
|------|--|----|
| 3.1  | Multiuser wireless communication system with a single base station   | 17 |
| 3.2  | Quantization regions and thresholds of the SNRs $\ldots$   | 19 |
| 3.3  | Uncertainty period   | 19 |
| 3.4  | Actual and Predicted SNR with prediction error for feedback SNR with period $w = 2$  | 23 |
| 3.5  | Average error variance (from left to right in this order) for different feedback<br>window. For each window, two groups of histograms are shown representing<br>IBF (left group) and PBF (right group)                         | 24 |
| 3.6  | Distribution of under, perfect, and over estimations (from left to right in this order) for different feedback window. For each window, two groups of histograms are shown representing IBF (left group) and PBF (right group) | 24 |
| 3.7  | Density distribution of prediction errors  | 27 |
| 3.8  | Numerical results for actual and predicted CQI   | 30 |
| 3.9  | Average BER for different feedback windows.  | 35 |
| 3.10 | Average BER of under estimated cases for different feedback windows  | 35 |
| 3.11 | Average BER of over estimated cases for different feedback windows   | 36 |
| 3.12 | Average throughput results for different values of $u$ (% of increase) and $v$ (% of decrease).  | 37 |
| 3.13 | MAC frame loss rate for different values of $u$ (% of increase) and $v$ (% of decrease).   | 37 |
| 3.14 | MAC frame retransmission rate for different values of $u$ (% of increase) and $v$ (% of decrease)  | 38 |
| 4.1  | OFDMA system   | 44 |
| 4.2  | FEREP Framework  | 46 |

| 4.3  | Overestimation in (a) interval based feedback (IBF) (b) prediction based feedback (PBF)                                 |    |  |
|------|---|----|--|
| 4.4  | System Goodput for available CQI slots with homogeneous users in (a) pedes-<br>trian mobility (b) vehicular mobility    |    |  |
| 4.5  | (a) CQI (b) Feedback Window for a MS with ITU vehicular B channel model   | 56 |  |
| 4.6  | Average feedback window for different $N_{ack}^{th}$ values with homogeneous users in vehicular mobility.               | 57 |  |
| 4.7  | System goodput for available CQI slots with heterogeneous users in (a) pedes-<br>trian mobility (b) vehicular mobility. |    |  |
| 4.8  | System goodput for available CQI slots with heterogeneous VoIP users in (a) pedestrian mobility (b) vehicular mobility  | 58 |  |
| 4.9  | System Goodput for available CQI slots with heterogeneous Video users in (a) pedestrian mobility (b) vehicular mobility |    |  |
| 4.10 | Average Feedback Window for available CQI slots in vehicular mobility with (a) VoIP users (b) Video users.              | 59 |  |
| 5.1  | The frame structure of the OFDMA/TDD system [29]  | 62 |  |
| 5.2  | Decision making for feedback scheduling strategy for user $k$   | 64 |  |
| 5.3  | Average feedback window for different vehicular mobility speeds   | 72 |  |
| 5.4  | Average feedback window for different $N_{ack}^{th}$ values   | 73 |  |
| 5.5  | Average throughput per user for varying application delay constraint  | 74 |  |
| 5.6  | Average packet loss rate per user for varying application delay constraint $\ldots$                                     | 74 |  |
| 5.7  | Average CQI feedback load for varying application delay constraint  | 75 |  |
| 5.8  | Average throughput per user for varying number of active users  | 75 |  |
| 5.9  | Average packet loss rate per user for varying number of active users $\ldots$ .   | 76 |  |
| 5.10 | Average CQI feedback load for varying number of active users  | 76 |  |

## List of Tables

| 3.1 | Minimum receiver SNR required to use in MCS   | 29 |
|-----|---|----|
| 3.2 | Simulation parameters   | 34 |
| 4.1 | Simulation parameters   | 53 |
| 4.2 | Minimum receiver SNR required to use in MCS   | 54 |
| 5.1 | Simulation Parameters   | 70 |
| 5.2 | Minimum receiver SNR required to use in MCS and their bitrates per symbol per carrier | 71 |

#### Acknowledgments

I would like to express my heartfelt gratitude to the many people who have supported me as I completed my graduate studies and dissertation.

First, I would like to sincerely thank my advisor, Dr. Lila Boukhatem, for providing continuous support and instructive guidance throughout my studies. I feel lucky to have an adviser like her with great patience and understanding to the difficult conditions that I had during this thesis. Her effort, encouragement and advices are greatly appreciated, and I will always be highly grateful to her for her commitment to my academic success. I present my sincere thanks to Prof. Khaldoun Al-Agha, for giving me the opportunity to work with his team and providing all the support to do my research.

I also owe a debt of gratitude to my colleagues of the Network Group in the Laboratory of the Research in Informatics (LRI), for their friendly welcome and helpful collaboration.

I wish to thank the reviewers, Prof. André-Luc Beylot and Dr. Thierry Turletti, to have the patience to read this dissertation and to give me valuable and detailed comments on my thesis. Thanks also to all the members of the jury, as it is a great honor for me to have them to evaluate my work.

I would like to emphasize my love and thanks to my family, especially my wife Sumaiya. Her understanding, encouragement and enduring love have sustained me throughout this journey toward completing my dissertation. Special thanks to my mother and father, for providing a loving and supportive environment throughout my childhood that fostered my academic success and for continued support today. I wish my father were alive to see this achievement of mine. My lovely son, Anjal, thank you because just your presence was enough to keep me on the right track.

xiv

### Acronyms

| ACK            | Acknowledgements                            |
|----------------|---|
| ADSL           | Asymmetric Digital Subscriber Line          |
| AR             | Autoregressive                              |
| ARQ            | Automatic Repeat Request                    |
| BER            | Bit Error Rate                              |
| BPSK           | Binary Phase Shift Keying                   |
| BS             | Base Station                                |
| $\mathbf{CDF}$ | Cumulative Distribution Function            |
| СР             | Cyclic Prefix                               |
| CQI            | Channel Quality Indicator                   |
| DCRA           | Dynamic Cqi Feedback Resource Allocation    |
| DFS            | Deterministic Feedback Scheduling           |
| DVB            | Digital Video Broadcasting                  |
| EMw-DCRA       | Dcra With Empirical Feedback Window         |
| FEREP          | Feedback Resource Allocation And Prediction |
| FH             | Frequency Hopping                           |
| FWA            | Feedback Window Adaptation                  |
| HARQ           | Hybrid ARQ                                  |
| HoL            | Head-Of-Line                                |
| IBF            | Interval Based Feedback                     |
| IFFT           | Inverse Fast Fourier Transform              |
| ISI            | Inter-Symbol Interference                   |
| LRP            | Long-Range Prediction                       |
| LTE            | Long Term Evolution                         |
| MARS           | Multivariate Adaptive Regression Splines    |
| MCS            | Modulation And Coding Scheme                |
| MMSE           | Minimum Mean-Square-Error                   |
| $\mathbf{MS}$  | Mobile Station                              |
| MUD            | Multiuser Diversity                         |
| NACK           | Negative Acknowledgement                    |
| NRT            | Non-Realtime                                |
| OFDM           | Orthogonal Frequency Division Multiplexing  |
|                |   |

| OFDMA           | Orthogonal Frequency Division Multiple Access    |
|-----------------|--|
| OFS             | Opportunistic Feedback Scheduling                |
| OIBF            | Opportunistic Interval Based Feedback            |
| OPBF            | Opportunistic Prediction Based Feedback          |
| PBF             | Prediction Based Feedback                        |
| PBFS            | Priority-Based Feedback Scheduling               |
| $\mathbf{PDF}$  | Probability Density Function                     |
| PER             | Packet Error Rate                                |
| PFS             | Proportional Fair Scheduling                     |
| PUSC            | Partially Used Subchannelization                 |
| $\mathbf{QAM}$  | Quadrature Amplitude Modulation                  |
| $\mathbf{QoS}$  | Quality Of Service                               |
| RLS             | Recursive Least Square                           |
| RT              | Realtime   |
| <b>RTw-DCRA</b> | Dcra With Realtime Feedback Window               |
| SMUD            | Selective Multiuser Diversity                    |
| SNR             | Signal To Noise Ratio                            |
| TDD             | Time Division Duplex                             |
| VoIP            | Voice-Over-Ip                                    |
| Wi-Fi           | Wireless-Fidelity                                |
| WiMAX           | Worldwide Inter-Operability For Microwave Access |
| WLAN            | Wireless Local Area Network                      |

### Chapter 1

## Introduction

Users today expect the same wireline broadband access and multimedia Internet experience from new generation of wireless networks. Recent 3G operators' statistics have shown a significant increase in mobile data usage with a rapid growth of traffic volume [82]. Many reasons are driving this trend: the proliferation of powerful smartphone devices, the increasing audio/video streaming and IPTV<sup>1</sup> demand and the diversified offers from operators for pricecutting flat-rate tariffs. To face this rapid growth in broadband data usage, most operators are installing the 4G networks that should provide a large range of new data and multimedia services with low costs, high reliability and quality. As an example, in USA, Sprint Nextel has begun using Mobile WiMAX<sup>2</sup> since 2008 and Verizon Wireless has began using LTE<sup>3</sup> service since 2010.

The high bandwidth and flexibility offered by 4G systems are mainly due to the orthogonal frequency division multiple access (OFDMA) technology which has been adopted in almost all wireless broadband access standards. OFDMA is an extension of orthogonal frequency division multiplexing (OFDM), which is currently the choice for high speed data access systems such as IEEE<sup>4</sup> 802.11n [88] wireless WLAN<sup>5</sup> (Wi-Fi<sup>6</sup>), IEEE 802.16e/m Mobile WiMAX [89] [90], 3GPP<sup>7</sup> LTE [1] and LTE-Advanced[108], etc. These systems are known as multi-user multi-carrier system where a base station can schedule more than one user at the same time using

<sup>&</sup>lt;sup>1</sup>Internet Protocol television

<sup>&</sup>lt;sup>2</sup>Worldwide Interoperability for Microwave Access

<sup>&</sup>lt;sup>3</sup>Long Term Evolution

<sup>&</sup>lt;sup>4</sup>Institute of Electrical and Electronics Engineers

<sup>&</sup>lt;sup>5</sup>Wireless Local Area Network

<sup>&</sup>lt;sup>6</sup>Wireless-Fidelity, name given to WLAN IEEE 802.11b standard-based products

<sup>&</sup>lt;sup>7</sup>3rd Generation Partnership Project

time-frequency resources.

#### **1.1** Motivation and Goals

In this thesis, we deal with the channel quality indicator (CQI) feedback reduction and resource allocation issues in multi-user multi-carrier OFDMA systems. Using CQI feedback information is common in wireless communications. This has been widely used for example in 2G/3G systems in the form of link adaptation. CQI feedback has become essential in 4G systems to have multi-user diversity, time-diversity, space diversity in  $MIMO^8$ , etc. As the wireless channel is dispersive by nature, each user needs to send the COI information of each subchannel to the base station to increase the reliability and achieve complete multiuser diversity gain. However, the available radio spectrum is scarce, and the energy consumption is limited. If the base station receives CQI feedbacks from all users, large feedback overhead exhausts the limited uplink capacity. The large feedback overhead issue can be solved by various feedback reduction schemes like opportunistic feedback, periodic feedback, best-m feedback. etc. However, opportunistic feedback strategy does not have any effect if the BS has enough capacity to schedule all the mobile stations in the same frame. The existing periodic feedback strategies do not take into account the over and under estimation of channels in between two consecutive feedbacks. Some periodic feedbacks consider the subcarrier power prediction in frequency or time domain to mitigate the feedback delay in the same physical frame but they do not consider long range prediction over multiple frames.

Considering above reasons, one of our goals is to introduce a periodic feedback mechanism which reduces the effect of over and underestimation of channels. We believe that there are real potentials to reduce the feedback overhead by means of CQI prediction using time domain correlations. Such approach brings new interesting issues including, e.g., performance analysis of the system in case of prediction errors and the mitigation of these prediction errors.

As the CQI feedback resource is limited, periodic feedback mechanisms allow each user to send its channel feedback once in every w frame (w is referred to as feedback window). A user with good channel does not need to send its feedback as frequently as an MS with bad channel. In such context, assigning a fixed w for each user is not optimal. However, 4th generation wireless systems are designed to support heterogeneous environments. On one hand, heterogeneity concerns the channel conditions experienced by users. On the other

<sup>&</sup>lt;sup>8</sup>Multiple Input Multiple Output

hand, the system is supposed to support heterogeneous applications, each of which has its specific requirement in terms of quality of service (QoS) [5]. Consequently, applying the same feedback strategy for all users is not adapted in these heterogeneous environments [60]. In such context, natural but crucial questions are raised: i) how to reduce the feedback overhead for better resource utilization without degrading the system performance, ii) how to allocate the available feedback slots in the OFDMA frames among the active heterogeneous MSs to send their channel condition, given the total budget constraint, and iii) how to design adapted feedback mechanism for heterogeneous MSs?

Having these questions in mind, another goal is to propose algorithms to adapt feedback window for each user dynamically, to determine users priority to send feedback according to their application profiles and QoS requirements. This approach brings other interesting issues including, e.g., varying feedback resource frame-by-frame for the system, determining feedback scheduling strategies for users with different QoS constraints, and identifying the link between feedback scheduling and downlink data scheduling.

#### **1.2** Contributions

Following the aforementioned challenges, the contributions of the present thesis can be summarized as follows.

- We propose a prediction based feedback (PBF) to reduce CQI feedback overhead by CQI prediction at the base station side. The results of this work has been published in [13]
  - M. A. Awal and L. Boukhatem, "Effect of Feedback Prediction on OFDMA System Throughput", International Wireless Communications and Mobile Computing Conference (IWCMC09), Leipzig, Germany, June 2009.
- We define the error distribution from the error data thus, we derive the probability of error in each adaptive modulation level.
- We present the analytical model for system performance with and without CQI prediction. The results of these works has been published in [14]
  - M. A. Awal and Lila Boukhatem, "Analysis of Feedback Prediction Error on the Downlink Performance of OFDMA Systems", IEEE Personal, Indoor and Mobile Radio Communications Symposium (PIMRC09), Tokyo, Japan, September 2009.

- We propose periodic feedback mechanisms named OIBF (opportunistic interval based feedback) and OPBF (opportunistic prediction based feedback) to mitigate the possible under and over estimated CQI. We exploit the hints from MAC<sup>9</sup> ARQ<sup>10</sup> protocol to enhance the performance of periodic feedback mechanisms. The results of this work has been published in [18]
  - M. A. Awal and Lila Boukhatem, "Opportunistic Periodic Feedback Mechanisms for OFDMA Systems under Feedback Budget Constraint", accepted to the IEEE Vehicular Technology Conference (VTC'11-Spring), Budapest, Hungary, May 2011.
- We propose feedback window adaptation (FWA) algorithm to determine the CQI feedback window frame-by-frame for each user exploiting cross-layer information. The results of this work has been published in [15]
  - M. A. Awal and Lila Boukhatem, "Cross-Layer CQI Feedback Window Adaptation for OFDMA Systems under Feedback Budget Constraint", accepted to the IEEE Symposium on Computers and Communications (ISCC'11), Kerkyra (Corfu), Greece, June 2011.
- We propose priority-based feedback scheduling (PBFS) algorithm with service differentiation in the case of total CQI feedback budget constraint.
- We propose a framework named FEREP (feedback resource allocation and prediction) to be used as plug-n-play in practical OFDMA systems without changing any parameters. The results of these works has been submitted for publication in [19]
  - M. A. Awal and Lila Boukhatem, "An Integrated Cross-Layer Framework of Adaptive Feedback Resource Allocation and Prediction for OFDMA Systems", submitted to Elsevier Journal of Computer Networks.
- We propose a dynamic CQI feedback resource allocation (DCRA) scheme which allocates the CQI resources dynamically on a frame-by-frame basis. The results of these works has been published in [16] [17]
  - M. A. Awal and Lila Boukhatem, "Dynamic CQI Resource Allocation for OFDMA Systems", accepted to the IEEE Wireless Communication and Networking Conference (WCNC'11), Quintana-Roo, Mexico, March 2011.

<sup>&</sup>lt;sup>9</sup>Medium Access Control

<sup>&</sup>lt;sup>10</sup>Automatic Repeat Request

- M. A. Awal and Lila Boukhatem, "Efficient CQI Resource Allocation for QoS Constrained OFDMA Systems", under preparation to be submitted in a journal.

Some initial works and literature review at the beginning of my PhD has been published in [12]

 M. A. Awal and L. Boukhatem, "WiMAX and end-to-end QoS support", Radio Resources Management in WiMAX: From theoretical capacity to system simulations, ISTE & Wiley, February 2009.

#### **1.3** Outline of the Thesis

The rest of this thesis is structured as follows.

- In chapter 2, we provide a literature review related to the feedback issues in OFDMA systems. At first, we introduce the multi-user multi-carrier OFDMA technology, then we present the CQI basics and main feedback mechanisms. We describe the related works dealing with feedback reduction and feedback resource allocation. Then we review the state of the art in channel predictions.
- Chapter 3 deals with reducing the feedback overhead using CQI prediction and reducing the over/under estimation effects in periodic feedback. At first, we introduce the system model we used in this thesis. Then, we show that it is useful to use channel prediction as a tool to reduce the feedback overhead and improve the uplink throughput. Then we introduce the analytical model for system performance with and without prediction. We define the error distribution from the error data to derive the probability of error in each adaptive modulation level and some performance metrics with both predicted CQI and actual CQI. In order to mitigate the possible under and over estimation effects of CQI prediction, we propose an opportunistic periodic feedback mechanism. In this mechanism, we exploit the hints from MAC-layer ARQ protocol to enhance the performance of periodic feedback mechanisms.
- Chapter 4 deals with adaptive feedback window, users priority, and limited feedback resource. As the fixed feedback window for all users is not optimal, we develop a CQI feedback window adaptation (FWA) algorithm to determine the feedback window for each user dynamically. The FWA algorithm uses cross-layer information like ARQ signals

in the absence of real CQI feedback. In current systems, the operators reserve a fixed number of slots for CQI feedback and this does not change frame-by-frame. If the number of users is greater than number of CQI slots, users compete for the CQI resources. We propose priority-based feedback scheduling (PBFS) algorithm to determine users priority to send feedback according to their service class and QoS requirements. We apply our prediction based feedback (PBF) to predict the CQI for the users not scheduled to send CQI feedback. We integrated all three modules under a framework named feedback resource allocation and prediction (FEREP).

- Chapter 5 deals with limited ARQ hints, feedback resource, and dynamic resource allocation. It is more natural to set smaller feedback window for that users experiencing heavily fluctuating channel so that BS knows their channel state more frequently. But because of fixed and limited number of CQI slots in each frame, the users miss the chance to send their CQI in time. To mitigate this issue, we propose to determine the number of CQI slots dynamically frame-by-frame. The ARQ protocol receives user acknowledgement only if the user is scheduled in the downlink. The reduction in users scheduling frequency also reduces the rate of ARQ hints and degrades the performance of above contributions. In this case, it is difficult to exploit the ARQ signal to adapt the feedback window for that user. In this chapter, we propose a dynamic feedback resource allocation (DCRA) scheme. DCRA uses two modes for feedback window estimation. The first one is an off-line mode based on empirical studies to derive optimal average feedback window based on user application and mobility profile. The second one is a realtime operation mode where the feedback window is dynamically adapted based on ARQ hints. Exploiting this feedback window information, DCRA allocates the CQI resources on a frame-by-frame basis to respect users QoS requirements.
- Chapter 6 concludes the document and summarizes the basic insights. Finally, it proposes suggestions for future research investigations.

### Chapter 2

## Literature Review

The background of the technologies and prior work related to the research of thesis are described in this section. Section 2.1 presents the OFDMA basics. Section 2.2 describes the CQI mechanisms and its importance in OFDMA based systems. Literature review of CQI feedback overhead is presented in Section 2.3 and literature of channel prediction is presented in Section 2.4.4.

#### 2.1 OFDMA Basics

Several of today's communication standards are based on orthogonal frequency division multiplexing (OFDM). In particular, OFDM is used in commercial standards for wireless local area networks (WLAN), namely, IEEE 802.11a [2]; IEEE 802.16 worldwide inter-operability for microwave access (WiMAX) standards [89] and the long term evolution (LTE) [1], LTE-Advanced [108]; for terrestrial digital video broadcasting (DVB-T) [40]; for terrestrial digital audio broadcasting (DAB-T) [39]; and for asymmetric digital subscriber line (ADSL) systems. Link adaptation, OFDM and multiple antenna methods are among the key physical layer solutions in most of the beyond 3G systems [105] like in LTE and the WiMAX standards.

OFDM [30, 84] is a multiplexing technique that subdivides the bandwidth into multiple frequency subcarriers. In an OFDM system, the input data stream is divided into several parallel sub-streams of reduced data rate (thus increased symbol duration) and each sub-stream is modulated and transmitted on a separate orthogonal subcarrier. The increased symbol duration improves the robustness of OFDM to delay spread. Furthermore, the introduction of



Figure 2.1: Frame structure of an OFDMA system (IEEE 802.16e)

the cyclic prefix (CP) can completely eliminate Inter-Symbol Interference (ISI) in case of CP duration is longer than the channel delay spread which eliminates the need for an equalizer. OFDM exploits the frequency diversity of the multipath channel by coding and interleaving the information across the subcarriers prior to transmissions. OFDM modulation can be realized with efficient Inverse Fast Fourier Transform (IFFT) with low complexity, which enables a large number of subcarriers (up to 2048). In an OFDM system, resources are available in the time domain by means of OFDM symbols and in the frequency domain by means of subcarriers as shown in Figure 2.1. The time and frequency resources can be organized into sub-channels for allocation to individual users. Orthogonal frequency division multiple access (OFDMA) is a multiple-access/multiplexing scheme that provides multiplexing operation of data streams from multiple users onto the downlink sub-channels and uplink multiple access by means of uplink sub-channels.

The OFDMA symbol structure consists of three types of subcarriers known as i) Data subcarriers for data transmission, ii) Pilot subcarriers for estimation and synchronization purposes and iii) Null subcarriers for no transmission; used for guard bands and DC carriers. Active (data and pilot) subcarriers are grouped into subsets of subcarriers called subchannels. OFDMA systems like WiMAX [89] and LTE-Advanced [108] supports sub-channelization in both downlink and uplink. There are two types of subcarrier permutations for subchannelization; diversity and contiguous. The diversity permutation draws subcarriers pseudo-randomly to form a sub-channel. It provides frequency diversity and inter-cell interference averaging. The contiguous permutation (also known as band-AMC) groups a block of contiguous subcarriers to form a subchannel. Contiguous permutation enables multi-user diversity by choosing the subchannel with the best frequency response. In general, diversity subcarrier permutations perform well in mobile applications while contiguous subcarrier permutations are well suited for fixed, portable, or low mobility environments. These options enable the system designer to trade-off mobility for throughput.

#### 2.2 Channel Quality Indicator (CQI)

In all the OFDMA systems, CQI at the receiver and/or transmitter is necessary for a number of advanced communication techniques. In particular, the transmitter needs CQI to apply link adaptation, pre-coding, pre-equalization, and adaptive transmit antenna diversity [56] [22] [115] [53]. Furthermore, at the receiver CQI is required for antenna combining and space-time decoding. For example, in [110] it has been found that CQI is important to realize the full potential of MIMO communication systems.

The mechanism behind the link adaptation is simple. If the channel is in a deep fade, a low order modulation is employed or the transmission is even truncated in this channel for a while. In the case of good channel conditions, a high data rate can be achieved employing high order modulation. Similarly, code rate and transmission power can be optimized according to the instantaneous channel conditions and the required transmission rate and reliability. Adaptive modulation and coding is a crucial part of all evolved communication standards. Goldsmith et al. has presented the channel capacity with channel side information for the different power and rate adaptation cases in [49]. Further results on spectral efficiency are shown in [48] [28] presenting the efficiency of power and rate adaptation also with a practical set of modulation alphabets. Adaptive modulation provides a significant performance enhancement also with fixed transmission power [28] [113].

In wireless communication systems, it is much more difficult to obtain reliable CQI than in wired systems. This is because the estimation error for time and frequency selective channels contains, in addition to a component due to noise. Schafhuber et al. presents different approaches to obtain the CQI in OFDMA systems [97]. The CQI feedback mechanism mainly depends on the subchannelization of OFDMA systems because of multipath fading and frequency selectivity. There are two types of sub-carrier permutations for sub-channelization

- Diversity Subchannelization
- Contiguous Subchannelization

When an MS is assigned a subchannel with diversity subchannelization (e.g. PUSC<sup>1</sup> in WiMAX), it sends the averaged SNR level of all the allocated subcarriers that are distributed over allowable spectrum range, to remove the effect of frequency-selective fading. In diversity, subcarriers in a subchannel are separated each other more than the coherence bandwidth and each subcarrier experiences independent rayleigh fading (multipath fading). Since a CQI feedback contains the averaged channel condition of the subcarriers, the effect of multipath fading on the CQI feedback will disappear due to averaging [123]. Therefore, there will be no effect of multipath fading and every subcarrier allocated to the MS uses the same modulation and coding scheme (MCS) level.

When an MS uses contiguous subchannelization (e.g. band-AMC in WiMAX), it sends the quantized signal to noise ratio (SNR) level of each subcarrier for each subband to a BS to utilize frequency-selective fading. According to CQI reporting method, the user may fully or partially utilize frequency-selective fading. If CQI is reported for each CQI band, then the SINR level of each subchannel can be known and thus frequency-selective fading can be fully utilized. CQI can be reported for some selected CQI band or only differential information can be reported [93]. In this case, the SNR level of each subchannel can be known partially and thus frequency-selective fading can be partially utilized [93] [26]. After the CQI is reported, the BS determines an MCS level for each subchannel on the basis of this information.

#### 2.3 CQI Feedback Overhead

Since the channel condition of a mobile station (MS) changes over time or frequency domain, due to the fading, interference and path loss, the recent wireless communication systems use link adaptation by means of adaptive modulation and coding scheme (MCS) and dynamic channel allocation. Adaptive MCS technique is used to the users data rate based on the

<sup>&</sup>lt;sup>1</sup>Partially Used Subchannelization

CQI in terms of signal-to-noise ratio (SNR). The adaptive MCS helps to increase the downlink spectral efficiency with the penalty of feedback overhead from each user and reduces the uplink efficiency for the users. A user needs to send  $N_{sc} \times b$  bits of information to the base station (BS) to inform full subchannel information, where  $N_{sc}$  is the number of subcarriers in a channel and  $b = log_2(\gamma_{sc})$  bits required to represent the SNR  $\gamma_{sc}$  for each subcarrier. To give a numerical example, for  $N_{sc} = 24$  as in WiMAX [89] with 10MHz spectrum bandwidth, the use needs to feedback at least 120 bits as CQI. Usually there are other additional information in a feedback message. So, the actual CQI feedback size is larger according to WiMAX [89].

The CQI feedback overhead problem becomes harder in case of using multiuser diversity provides a further degree-of-freedom that can be utilized in multiuser communication systems. When the number of users is large, there is a high probability that at least one of the users has a strong channel [72]. The basic principle to maximize multiuser diversity is to transmit to the users with the largest receive signal-to-noise ratio (SNR) at each channel resulting in maximum system throughout [110][72].

Another area of research experiences tremendous CQI feedback overhead is dynamic channel allocation. Dynamic channel assignment for multiple users can be performed in the time [72], frequency [115] and space [110] domains for both the uplink and downlink directions. To know the full channel information, each user needs to inform the CQI for all the subchannels. So, the feedback over increases to  $N_{sc} \times log_2(\gamma_{sc}) \times N_{ch}$  bits per user where  $N_{ch}$  is the number of subchannels in the system.

#### 2.4 CQI Feedback Reduction

#### 2.4.1 Quantization

A straightforward way to decrease the feedback rate is by quantization of the SNR measurements before feedback transmission. Since the SNR values are real-valued, some quantization must be applied; the question is how few bits can be used with a maintained high total throughput of the multiuser system. Johansson [66] and Floren et al. [44] are early studies of the effect of feedback quantization. They conclude that 1-bit quantization is good enough in many cases, if the average SNR of each user is known. In their example, a throughput of more than 90% of the throughput with unquantized feedback is achievable, with a 1-bit quantizer optimized for the correct average user SNR and number of users. Sanayei et al. [96] and Somekh et al. [102] shows that one bit feedback per user can provide nearly maximum sum-rate capacity with a properly defined threshold when the number of users is large. But these works are for TDMA where only user is served in each scheduling slot out of large number of active users. The OFDMA system like Mobile WiMAX uses 5-bits quantization for.

#### 2.4.2 Opportunistic Feedback

The feedback load of multiuser systems can be reduced by using selective multiuser diversity (SMUD) [46] [47] [54] where only the users enjoying a good channel state send a feedback message. In [46], Gesbert and Alouini propose a SMUD scheme, by letting the users report CQI only when the channel quality exceeds a predefined threshold. If no users have an SNR that exceeds the threshold, a random user is scheduled. When the number of users is large K > 25, the SMUD scheme leads to a remarkable reduction of the average number of users transmitting feedback, with a reduction of the feedback rate down to 10% of full feedback for the best case, with small or no loss in terms of system capacity. Hassel et al. [55], improves the SMUD scheme by requesting full feedback transmission from all users if no user had a SNR above threshold; this guarantees the full multiuser diversity gain, at the cost of some extra feedback compared with original SMUD. These opportunistic mechanisms behaves like a max-SNR scheduling scheme by letting the users with highest channel quality to send the feedback. To maintain the fairness among users, Yang et al. [118] propose to select the user k with the largest  $\gamma_k/\bar{\gamma_k}$ , where  $\gamma_k$  is the CQI of user k, and  $\bar{\gamma_k}$  is the short-term averaged CQI.

Hassel et al. [54] develops the SMUD scheme further by exploiting multiple SNR thresholds used sequentially. The base station first requests feedback from those users with an SNR exceeding the highest threshold. If there are none, the threshold is successively lowered until a user fulfilling the requirement is found. Special attention is given to problems with feedback collisions, when several users reply to the same threshold query, and to scheduling outage with a scheduling deadline, when no feedback is received before a scheduling decision must be made. However, opportunistic feedback originally was proposed for TDMA or FDMA system. The feedback messages of different users may collide in the feedback channel due to the contention based feedback transmission, resulting in a loss of system performance. However, the bandwidth of the feedback channel is limited despite of the number of users. Opportunistic feedback provides a good tradeoff between the system performance and the feedback overhead [46]. Multiuser diversity and threshold optimization with an imperfect estimator and a noisy feedback channel have been addressed in [117]. The drawback to the subchannel-wise feedback method is the fact that the feedback overhead grows linearly with the increasing number of subchannels. The feedback overhead can be reduced by using a group-wise one bit strategy, in which one bit indicates the quality of the group of channels [24] [25]. Opportunistic feedback has been examined also for OFDMA systems in several studies [4] [25] [24] [6] [3]. The authors in [4] [6] [3] have studied weighted sum rate maximization for sequential and contention based feedback schemes.

#### 2.4.3 Best-M Feedback

Svedman et al. [106] [91] discuss an OFDM multiuser scenario, where each user sends feedback about the *M* best subcarriers. To reduce the feedback rate, they divide the subcarriers into subchannels, and each user computes the average SNR within the subchannels and feeds back the indices and CQIs of the *M* strongest subchannels. For each subchannel, the base station chooses the user with the highest CQI. With this scheme, some subchannel indices may not be fed back by any user, and then the available transmit power is divided only among the assigned subchannels. The best-M methods have been considered also for practical networks such as WiBro [26]. Moreover, the total feedback load can be controlled using an adaptive and selective feedback scheme which takes the number of users, fairness and the QoS into account [106]. A QoS aware best-M feedback method has been considered also for MIMO systems [27]. The drawback of the best-M feedback scheme is the fact that it is sensitive to feedback bit errors [91]. For that reason, Kovacs et al. [74] use a bit mask to indicate the M best channels resulting in the same overhead as in the RB-wise one bit feedback scheme. Because of feedback budget constraint and increase of high datarate applications, practical systems uses best-1 feedback.

#### 2.4.4 Channel Prediction

When the CQI feedback arrives to the transmitter, the estimated CQIs are outdated after a time period equal to a fraction of the channel coherence time. Therefore, to obtain up-to-date CQI, time and frequency selective channels need to be tracked continuously. For techniques that require CQI at the transmitter, outdated CQI is a severe problem. If CQI is obtained from the receiver via a feedback link, a significant percentage of the data rate of the feedback link may be required to transmit channel parameters. Here, CQI may be outdated due to transmission delays. On the other hand, in a time division duplex (TDD) communication

scheme, if the channel is estimated by the transmitter while in receive mode, this CQI could be outdated as well when applied subsequently. Depending on the application, accurate CQI is required to achieve performance gains similar to those that have been demonstrated with perfect channel knowledge. Communications over time and frequency selective fading channels inherently suffers from channel uncertainty. The detrimental effects of channel uncertainty can be particularly pronounced for large bandwidths.

Recently, several techniques were developed to predict future behavior of the mobile channel. In [35], the multivariate adaptive regression splines (MARS) model was used to capture the dynamics for predicting parameters of wideband fading channels several millisecond ahead for fast vehicle speed. The sub-space based [57] root-MUSIC method [61] and ESPRIT type algorithm [10] were employed to estimate the power spectrum that constitutes the fading process. Then these sinusoids were extrapolated to predict future samples. These methods were tested using synthetic and measured data, and it was concluded that reliable prediction is feasible at least one wavelength into the future.

An adaptive long-range prediction (LRP) method for flat fading channel was also proposed in [41] [42] [59] [34]. This algorithm employs an autoregressive (AR) model to characterize the fading channel and computes the minimum mean-square-error (MMSE) estimate of a future fading coefficient based on a number of past observations. The advantage of this algorithm relative to conventional methods is due to its low sampling rate (on the order of twice the maximum Doppler shift and much lower than the data rate), which results in longer memory span and prediction further into the future for a fixed filter length. The low sampling rate also results in reduced feedback rate. The LRP can be implemented adaptively and thus is less complex and more robust than other fading prediction techniques. More recently, the LRP was extended to frequency selective channels. In [119] [120], LRP and adaptive modulation using CQI of another carrier was addressed, and in [77], LRP for frequency hopping (FH) systems was investigated. In [33], they developed the long-range prediction algorithms for OFDM systems. An ideal MMSE method that utilizes previous observations in time and frequency domain, and robust adaptive LRP algorithms are developed and compared. The LRP is utilized in adaptive bit and power allocation for the OFDM system.

As the channel statistics are unknown in practical applications, and thus they would have to be estimated prior to the design of the MMSE-based channel predictor [97]. The pilot symbol assisted MMSE channel estimation is shown problematic since practical channels are stationary only for a certain time. Fortunately, the estimation of the channel statistics can be completely avoided by the application of adaptive channel predictors like normalized leastmean-square (NLMS) algorithm or the recursive least-squares (RLS) algorithm. The adaptive channel predictors do not require any statistical prior knowledge and they are able to track nonstationary channel and noise statistics. We consider RLS-based prediction in this thesis as they converges faster compared to the NLMS algorithm and and has a smaller excess MSE [97].

Ekman et al. [36] deal with the channel prediction with single carrier systems like GSM. They showed the noise reductions in channel estimation and then showed the prediction of the power. They applied the channel prediction in link adaptation to increase the system performance. For the end user point of view, the channel power or channel tap prediction complexity increase with number of subcarriers. To apply the idea of RLS-based prediction to predict the CQI feedback of OFDMA systems, we also propose to implement the prediction algorithm in BS-side to reduce the MS-side complexity. The use of RLS-based prediction algorithm has also been proposed in [68] [33]. Jordan et al. [68] used RLS in opportunistic beamforming to mitigate the reporting delay from the MS because of propagation delay. They calculate filter coefficients only once for every angle in the beamforming process and reuse these coefficients in further cyclic periods of the beamforming process. Duel-Hallen et al. [33] used RLS-based long range prediction to track correlated sub-carriers so that the MS sends only one CQI for all the correlated sub-carriers.

Adaptive modulation was combined with single carrier channel prediction also shown in [59] [58] [34] [43] [87]. Adaptive OFDM with imperfect channel state information was analyzed in [103] [116] [122].

### Chapter 3

# Prediction Based CQI Feedback To Reduce Feedback Overhead

CQI prediction has been introduced in many works [36] [97] [33] [7] to reduce the CQI feedback overhead. Most of the feedback prediction proposals are aimed at mitigating the feedback reporting delay and maximizing the downlink throughput. A few of them [36] proposed to predict the CQI several frames in advance, but in fact, they predicted only one frame in advance. Ekman et al. shown in [36] that the channel prediction provides significant gain in in adaptive MCS upto 10ms which is equivalent to 2 frames duration for practical system like Mobile WiMAX. Besides, reducing the uncertainty effect of channel prediction has received much less attention. Our goal is to reduce the CQI feedback overhead as much as possible related to each user by predicting the CQI in the BS and use this saved bandwidth to carry more user data traffic, hence, increase the overall uplink throughput.

This chapter shows how the CQI prediction can provide significant gain in feedback reduction with negligible degradation of the downlink throughput. We present our total overhead reduction model in two phases. In first phase, we use the RLS<sup>1</sup>-based time-domain CQI prediction and investigate its feasibility in multiuser environments. Like any other prediction algorithm, RLS generates prediction errors and it increases with prediction horizon. Then, we determine the effect of the prediction error on the downlink performance. In a second phase, we introduce a cross-layer based opportunistic periodic feedback algorithm to reduce the effect of prediction errors. The BS takes hints from ARQ protocol about the possible prediction error and creates a vector of scaling coefficients. Then it applies these scaling coefficients in

<sup>&</sup>lt;sup>1</sup>Recursive Least Square

the past history of CQI values to mitigate the prediction error in future predicted CQI.

#### 3.1 System Model



Figure 3.1: Multiuser wireless communication system with a single base station

We consider an OFDMA cellular system with bandwidth B consisting of  $N_c$  subcarriers. The system has a set  $\mathcal{K}$  of K simultaneously active mobile stations (MSs, also referred to as users throughout the paper) communicating with a single base station (BS) as shown in Figure 3.1. The subcarriers are distributed among the users using partially used subchannelization (PUSC) and each user subchannel has  $N_s = N_c/K$  subcarriers. The BS and all MSs have one antenna each. The channel process of each user is assumed independent and stationary. The channel gain is assumed constant over a frame duration  $T_f$ , but may vary frame-by-frame. The signal received by user k at frame t, is given by

$$y_k(t) = H_k(t)x_k(t) + n_k(t), \quad k = 1, 2, ..., K$$
(3.1)

where  $x_k(t) \in \mathbb{C}^{N_s \times 1}$  is the complex transmitted signal,  $y_k(t) \in \mathbb{C}^{N_s \times 1}$  is the complex received signal, and  $n_k(t) \in \mathbb{C}^{N_s \times 1}$  is assumed to be a zero mean complex Gaussian noise vector with variance  $\sigma_n^2$ ; and  $H_k(t)$  is the diagonal channel response matrix given by  $H_k(t) =$  $diag\{h_{k,1}(t), ..., h_{k,N_s}(t)\}$ , where  $h_{k,n}(t)$  are the complex valued wireless channel fading random processes and  $h_k(t) \sim \mathcal{CN}(0_N, \sigma_h^2)$  is i.i.d over different users. The instantaneous Signal-to-Noise Ratio (SNR) of subcarrier *i* in frame *t* for MS *k* is defined as  $\gamma_k^i(t) = \frac{|h_{k,i}(t)|^2}{\sigma_n^2}$ .
### 3.1.1 CQI Measurements

There are several works discussed in literature about using a single value to denote the CQI for a subchannel or a group of subcarriers [114]. Each MS k measures average SNR  $\gamma_k^{avg}$  over all the subcarrier of the preamble except the guard subcarriers and the DC subcarriers as below.

$$\gamma_k^{avg} = \frac{1}{N_c} \sum_{i=1}^{N_c} \gamma_k^i \quad \text{(in dB)}$$
(3.2)

where  $N_c$  is the number of subcarriers except the guard and DC subcarriers, and  $\gamma_k^i$  is estimated SNR of user k on subcarrier i. Reporting this SNR would entail perfect CQI at the BS. To provide such high precision for CQI, the number of bits would be large. Since feedback load of the SNR quantization increases rapidly with the number of quantization levels, onebit quantzation has been studied in [96] [24]. Optimal distribution of feedback bits with quantization precision is discussed in [6]. We assume to use the CQI statistics be quantized in 1 dB increments according to IEEE 802.16e [89] shown below

$$\gamma_k^{avg} \approx \left\lceil \frac{1}{N_c} \sum_{i=1}^{N_c} \gamma_k^i \right\rceil \quad (\text{in dB})$$
(3.3)

It is assumed that the BS receives SNR without any error and with zero-delay from the MS.

We note that in case of contiguous subchannelization like band-AMC, averaged CQI does not provide the BS with any knowledge on the frequency selectivity. To cope with this channel variations in frequency selectivity, effective SNR is also adopted [121] in mobile WiMAX. Effective SNR is defined as

$$\gamma_k^{eff} \triangleq -\beta ln(\frac{1}{N_s} \sum_{i=1}^{N_s} e^{\frac{\gamma_k^i}{\beta}}) \quad \text{(in dB)}, \tag{3.4}$$

where  $\gamma_k^i$  are the per subcarrier SNR values of user k and which are typically different in a frequency selective channel. Parameter  $\beta$  is function of MCS for a given coding scheme. MS reports the effective SNR to the BS, and allows the BS to decide MCS level and power boosting adjustment. In contrast to the averaged CQI in equation 3.2, the power adaptation for each effective SNR is MCS dependent and does not change linearly. For simplicity, in this study, we used the averaged CQI under PUSC assumptions.

### 3.1.2 Adaptive Modulation and Coding Scheme

The system uses MCS, that is, the coding scheme and the modulation constellation depend on the channel SNR of the selected user [28]. The transmit power is allocated to each subchan-



Figure 3.2: Quantization regions and thresholds of the SNRs

nel identically. Upon receiving the SNR from the MS, the BS maps it into a quantization level. Let  $\{Q_j\}_{j=0}^{m-1}$  be the *m* quantization levels (Fig. 2) of the SNRs defined by  $Q_j = [q_j, q_{j+1})$ , with  $0 \leq j \leq m-1$ ,  $q_0 = 0$  and  $q_m = \infty$ , where  $q_j$  are the SNR values to limit each quantization level. Specifically, we assign one MCS level to each quantization level  $Q_j$  and a corresponding data rate of  $r_j$  bits/symbol. When the instantaneous SNR  $\gamma$  falls within a given quantization level, the associated signal constellation is transmitted. No signal is transmitted if  $\gamma \leq q_0$ . The best quantization varies with the mean SNR, the number of available resources, the user position in the queue and the selected optimization criteria [78]. Moreover, the lower and upper boundaries for each MCS level can be selected according to the performance of the available modulation methods and coding rates with a predefined error rate constraint [8]. This thesis does not cover the optimization of the SNR quantization regions but related literature can be found in [8] [9].



#### 3.1.3 Uncertainty Period

Figure 3.3: Uncertainty period

We define the uncertainty period as the period when the BS does not have any knowledge

about any possible channel change. We illustrate the uncertainty period by an example shown in Figure 3.3 where the BS receives CQI feedback from an MS at frames t and  $t + 4T_f$ . The time duration during the frames  $t + T_f$  to  $t + 3T_f$  is the uncertainty period when the BS does not receive any feedback.

### 3.1.4 Automatic Repeat Request

The system uses automatic repeat request (ARQ) in the data link layer to increase the reliability of packets. During a packet transmission, the MS uses explicit acknowledgements (ack) and negative acknowledgement (nack) packets to inform BS that it has received the packet correctly or not. The BS sends the next packet on the queue if it receives an ack from the MS. The BS retransmits the last sent packet if it receives a nack message from the MS. The sender may also uses implicit nack with the help of a timer. Upon sending a packet, the BS starts a timer. If a specific period of time elapses before an ack is received (which is termed as *timeout*), the BS assumes that the packet is not received correctly and retransmits the packet. The BS usually re-transmits the packet. If the maximum number of retransmission  $N_{re}^{max}$  for each packet is very large, it would violate the delay constraints for real-time applications like VoIP, video streaming, etc. Liu et al. [79] proposes truncated ARQ combined with MCS to limit the  $N_{re}^{max}$  according to the delay constraints of real-time applications.

### 3.1.5 Doppler Frequency

A user channel quality is subject to various fading effects<sup>2</sup> (environment, fading, mobility, etc.). Doppler frequency is used to represent the user mobility which is denoted by  $f_d$ ,  $f_d = s/\lambda$ , and  $\lambda = c/f_c$ . s is user speed, c is the speed of light, and  $f_c$  is the carrier frequency. We will use this Doppler frequency information in later chapters of this thesis.

### **3.2** Interval Based Feedback (IBF)

Interval based feedback has been used in many works [86] [64] [89] [112] [31]. In this technique, each MS is allowed to send its channel feedback periodically once in every w frame. Once a feedback is received, BS acts on deciding the MCS level to use for next w frames

<sup>&</sup>lt;sup>2</sup>Since we assume a single OFDMA cell, inter-cell interference is not considered.

to transmit till the next feedback is received and regardless of probable change in channel condition within this w feedback period. Here, the BS sends data to a MS using last determined MCS level based on last received CQI. If the BS receives a CQI at frame t, the amount of error  $e_{ibf}$  caused during the uncertainty period can be measured by

$$e_{ibf}(t+i) \triangleq \gamma(t) - \hat{\gamma}(t+i), \quad 1 \le i < w - 1 \tag{3.5}$$

where  $\hat{\gamma}(t+i)$  is the estimated CQI at the *i*th frame of the uncertainty period. Let  $Q_j$  be the selected MCS level for last received CQI  $\gamma_{last}$ . For any  $e_{ibf} > 0$ , if the current MCS level should be  $Q_{j-1}$  but the BS still using MCS level  $Q_j$ , the situation is considered as overestimated. Overestimated CQI introduces an increase in the system throughput while increasing the BER as well, hence increases the chance of packet loss. On the other hand, for any  $e_{ibf} < 0$ , if the current MCS level should be  $Q_{j+1}$  but data are still being sent using MCS level  $Q_j$ , the situation is considered as underestimated. Underestimated CQI leads to a decrease in the system throughput but also reduces the packet lossrate because of lower BER. Perfect MCS estimations occur either if no error is made or the induced error maintains  $\hat{\gamma}(t+i)$  in MCS level  $Q_j$ .

### **3.3** Prediction Based Feedback (PBF)

In this section, we discuss our prediction based feedback mechanism. We first describe our CQI prediction model. Then we describe the model to determine the prediction error probability and derive the system performance.

#### 3.3.1 Prediction Model

We propose to implement a RLS prediction algorithm [94] in the BS-side for the specific purpose of anticipating the future SNR level of the received signal at the receiver side. In our work, we assume that each MS sends the average SNR for all the sub-carriers assigned to it. We use short range prediction in every frame to estimate the most probable SNR level that the receiver is going to experience and avoid explicit feedback from the receiver. The predictability relies on error measures expressed in terms of a time average of the actual received SNR instead of a statistical average. The least square error  $e_{RLS}$  for a MS based on time average of window size w is defined as [94]

$$e_{rls}(t) = \sum_{j=t-w}^{t} (\lambda^{(t-j)} e^*(t) e(t))$$
(3.6)

where  $\lambda$  is the scalar weighting factor with  $0 < \lambda \leq 1$  that can change the performance of the prediction.  $\lambda$  is applied to the previous w input data.  $e^*(t)$  is the complex conjugate of the prediction error e(t). At any frame t, the amount of error  $e_{pbf}$  caused by prediction can be measured by

$$e_{pbf}(t) \triangleq \gamma(t) - \hat{\gamma}(t), \quad t \ge w$$
(3.7)

where  $\gamma(t)$  is the actual SNR experienced by the MS during frame t, and  $\hat{\gamma}(t)$  is the predicted SNR for frame t. The SNR is predicted as [94]

$$\hat{\gamma}(t) \triangleq \mathbf{Z}(t-1)\Gamma_{pbf}(t), \quad t \ge w$$
(3.8)

where  $\hat{\gamma}(t)$  is the predicted SNR for frame t, Z(t) is the w-th order prediction filter coefficient and  $\Gamma_{pbf}(t)$  is the previous real or predicted SNRs of moving window of size w up to frame t. The prediction filter Z(t) is determined using RLS gain vector G and covariance matrix M. The prediction filter coefficient is updated as

$$Z(t) = Z(t-1) + G(t)e_{pbf}^{*}(t), t \ge 1$$
(3.9)

where  $e_{pbf}^*$  is the conjugate of prediction error  $e_{pbf}$ . Vector G is calculated recursively as

$$\mathbf{G}(t) \triangleq \frac{\mathbf{M}^{-1}(t-1)\Gamma_{pbf}(t)}{\lambda + \Gamma_{pbf}^{T}(t)\mathbf{M}^{-1}(t-1)\Gamma_{pbf}(t)}, t \ge w$$
(3.10)

Here we observe that the inverse of M is required. To reduce the time complexity, the inverse of M can be recursively calculated as

$$\mathbf{M}^{-1}(t) \triangleq \frac{1}{\lambda} [\mathbf{M}^{-1}(t-1) - \mathbf{G}(t) \Gamma_{pbf}^{T}(t) \mathbf{M}^{-1}(t-1)], t \ge w$$
(3.11)

where  $\Gamma_{pbf}^{T}(t)$  is the transpose of  $w \times 1$  matrix  $\Gamma_{pbf}(t)$  and  $\lambda$  is the scalar weighting factor with  $0 < \lambda \leq 1$  that can change the performance of the prediction. The recursion is initialized as

$$Z(0) = G(0) = \gamma(0) = 0, M^{-1}(0) = dI_{wu}$$

where  $I_{ww}$  is an  $w \times w$  identity matrix and d is a large positive constant. We put d = 5000. We summarize the whole algorithm in Algorithm 1

To show the performance of CQI prediction using RLS algorithms, we ran simulations for users having rayleigh channel with ITU pedestrian (3km/h) and ITU vehicular mobility

| Algorithm 1 CQI Prediction Algorithm  |                       |
|---|-----------------------|
| 1: procedure PBF  |                       |
| 2: input: $\gamma(t),  \forall t \in \mathbb{N}$  |                       |
| 3: <b>output:</b> $\hat{\gamma}_{opbf}(t),  t \geq w \text{ and } \forall t \in \mathbb{N}$   |                       |
| 4: initialize: $\delta = 5000, \lambda = 0.99, Z(0) = 0, G(0) = 0, M^{-1}(0) = \delta I_{ww}$ |                       |
| 5: for each scheduling time, $t = w, w + 1,$ compute do                                       |                       |
| 6: $\hat{\gamma}_{pbf}(t) = \mathbf{Z}(t-1)\Gamma_{pbf}(t)$                                   | $(w^2 \text{ flops})$ |
| 7: $e_{pbf}(t) = \gamma(t) - \hat{\gamma}_{pbf}(t)$   | (w  flops)            |
| 8: $\alpha = \Gamma_{pbf}^T(t) \mathbf{M}^{-1}(t-1)$  | $(w^2 \text{ flops})$ |
| 9: $\beta = \lambda + \alpha \Gamma_{pbf}(t)$   | (w  flops)            |
| 10: $G(t) = \frac{\alpha^T}{\beta}$   | (w  flops)            |
| 11: $M' = G(t)\alpha$   | $(w^2 \text{ flops})$ |
| 12: $M^{-1}(t) = \frac{1}{\lambda} [M^{-1}(t-1) - M']$  | $(w^2 \text{ flops})$ |
| 13: $Z(t) = Z(t-1) + M(t)e(t)$  | (w  flops)            |
| 14: end for   |                       |



Figure 3.4: Actual and Predicted SNR with prediction error for feedback SNR with period w = 2.

(3km/h). Figure 3.4 shows the performance of real SNR  $\gamma_k$ , predicted SNR  $\hat{\gamma}_k$  of user k with period w = 2 and the error according to Eq. 3.7. The value w means the MS sends one CQI feedback out of w frames. In other words, BS predicts the CQI for w - 1 frames in every w frames. Here, for simplicity and visibility, we show the result of first 100 frames only. The prediction error seems too high at the initial iterations of the algorithm as the covariance knowledge of the CQI feedback is zero. With further iterations, the algorithm gains knowledge about the covariance and reduces the prediction error. We observe that the prediction error of RLS algorithm converges as early as 7 frames (Figure 3.4).



Figure 3.5: Average error variance (from left to right in this order) for different feedback window. For each window, two groups of histograms are shown representing IBF (left group) and PBF (right group).



Figure 3.6: Distribution of under, perfect, and over estimations (from left to right in this order) for different feedback window. For each window, two groups of histograms are shown representing IBF (left group) and PBF (right group).

In Figure 3.5, we compare the estimation error variance of IBF and PBF schemes according to Eq. 3.5 and Eq. 3.7, respectively. It is trivial that lower mobility case shows smaller error variance for both methods. On the other hand, IBF method always presents larger error variance than PBF method. For high mobility case, the difference between error variance increases with the increase in feedback window.

In Figure 3.6, we compare the MCS decision distribution for both IBF and PBF methods. The MCS decisions fall in three categories which are under, perfect and over estimations. Both feedback methods lose the ability of perfect estimation with mobility and feedback window. In IBF method, the under and over estimations are symmetric meaning the rate of under and over estimations are equal. In PBF scheme, the under estimation rate is higher than the over estimation rate in all cases. This can be explained by the behavior of our RLS algorithm which is biased towards the underestimation. It is possible to create an algorithm biased towards the overestimation or unbiased but that will increase the BER and increase packet loss rate. The biased behavior of the RLS algorithm is detailed in [36].

### 3.3.2 Effect of Feedback Prediction Error

In this section we propose an analytical procedure to investigate the effect of feedback prediction on downlink system performance. Based on the collected prediction error data from simulation experiments we obtained using the Algorithm 1, we define the error distribution. Then, we derive the probability of error in each MCS level. We define three performance parameters which enables us to obtain the average spectral efficiency, average bit error rate (BER) and average throughput. These metrics will be evaluated with both predicted feedback and actual feedback. For brevity, we conduct the analysis in this thesis based on the error distribution obtained for w = 2. Similar analysis can be done for the error distributions derived for higher feedback windows (w > 2).

Let us define the probability density function (PDF) and cumulative distribution function (CDF) of SNR  $\gamma$  are denoted  $p_{\gamma}(\gamma)$  and  $P_{\gamma}(\gamma)$ , respectively. Due to the Rayleigh fading assumption  $\gamma$  is exponentially distributed with mean  $\bar{\gamma}$  as below [92].

$$p_{\gamma}(\gamma) = \begin{cases} \frac{1}{\bar{\gamma}} e^{-\gamma/\bar{\gamma}} & \gamma \ge 0\\ 0 & \text{otherwise} \end{cases}$$
(3.12)

$$P_{\gamma}(\gamma) = 1 - e^{-\gamma/\bar{\gamma}} \quad \gamma \ge 0 \tag{3.13}$$

The spectral efficiency of our modulation scheme equals its average data rate per unit bandwidth (D/B). When we send  $r_j$  bits/symbol for some SNR  $\gamma \in Q_j$ , the instantaneous data rate is  $r_j/T_s$  (bps), where  $T_s$  is the symbol time. Assuming Nyquist data pulses  $(B = 1/T_s)$ , for discrete rate adaptation, the spectral efficiency is given by

$$\frac{D}{B} = \sum_{j=0}^{N-1} r_j \int_{q_j}^{q_{j+1}} p_\gamma(\gamma) d\gamma \qquad \text{bps/Hz}$$
(3.14)

and the throughput is given by

$$\eta = \sum_{j=0}^{N-1} r_j \int_{q_j}^{q_{j+1}} (1 - \text{BER}(\gamma))^L p_\gamma(\gamma) d\gamma \qquad \text{bps}$$
(3.15)

where L stands for the number of bits per transmission to one user. As shown by Goldsmith et al. [28], the BER for M-QAM modulation schemes (and also for BPSK) can be approximated for each user with unit power as

$$BER(\gamma) \approx 0.2e^{-1.6\frac{\gamma}{2^{r(\gamma)}-1}} \tag{3.16}$$

where the  $r(\gamma) = \lfloor (\log_2(\gamma) \rfloor$  is the corresponding bits/symbol for  $\gamma$  and  $\lfloor x \rfloor$  is floor operation which provides largest integer not greater than x. Hence the average BER is computed as

$$\overline{\text{BER}} = \frac{E[\text{number of error bits per transmission}]}{E[\text{number of bits per transmission}]} \\ = \frac{\sum_{j=0}^{N-1} r_j \int_{q_j}^{q_{j+1}} BER(\gamma) p_{\gamma}(\gamma) d\gamma}{\sum_{j=0}^{N-1} r_j \int_{q_j}^{q_{j+1}} p_{\gamma}(\gamma) d\gamma}$$
(3.17)

Using 3.13, the probability that  $\gamma$  lies in quantization level j before applying the prediction, i.e.,  $P(\gamma \in Q_j)$  is given by

$$P(\gamma \in Q_j) = P_{\gamma}(q_{j+1}) - P_{\gamma}(q_j)$$
  
= 
$$P(\xi|\xi=0) \int_{q_j}^{q_{j+1}} p_{\gamma}(\gamma) d\gamma \qquad (3.18)$$

where  $P(\xi)$  denotes the probability of prediction error  $\xi$  and  $P(\xi|\xi = 0) = 1$  as there is no prediction applied yet.

Let denote the PDF of the prediction error  $\xi$  as  $p_{\xi}(\xi)$  for  $-\infty \leq \xi \leq \infty$ . Fig. 3.7 shows the density distribution of all the prediction errors obtained in from our previous experiments with 30 users and 10 seeds. The distribution shows a very high pick around 0 and comparatively much smaller picks around  $\pm 1$  and  $\pm 2$ . The distribution does not fit to any standard



Figure 3.7: Density distribution of prediction errors

distribution because of these features. Consequently, we used kernel density estimation [98] to fit our error distribution as shown in Fig. 3.7. The error estimation can then be defined as

$$p_{\xi}(\xi) = \frac{1}{Hs} \sum_{i=1}^{H} \mathcal{N}(\frac{\xi - \xi_i}{s})$$
(3.19)

where  $\mathcal{N}$  is a kernel function and s is a smoothing parameter called the bandwidth.  $\{\xi_i\}_{i=1}^H$  are collection of H histograms identically distributed samples of a random variable. The kernel  $\mathcal{N}$  is taken to be a standard gaussian function with mean zero and variance 1. Thus the variance is controlled indirectly [98] through the parameter s as

$$\mathcal{N}(\frac{\xi - \xi_i}{s}) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(\xi - \xi_i)^2}{2s^2}}$$
(3.20)

Due to the tail area of the error distribution, the prediction algorithm will provide the underestimated or overestimated SNR compared to the actual SNR. In case of  $P(\xi < 0)$ , this underestimated SNR could act on deciding lower MCS level and results in throughput degradation. At the same time, while the actual MCS should be higher, choosing the lower level of MCS would act on reducing the BER. Similarly, in case of  $P(\xi > 0)$ , the overestimated SNR could act on deciding higher MCS level and results in throughput enhancement with the cost of an increased BER.

To determine the effect of underestimation or overestimation, we first need to determine the probability of predicted SNR falling in reference quantization level  $Q_j$ . This probability depends on three parameters:

- the width of that reference quantization level (i.e.  $q_{j+1} q_j$ )
- how far the actual instantaneous SNR value is from the boundaries of quantization level  $Q_j$ ?
- the amount of prediction error  $\xi$

The condition  $\{(\hat{\gamma} \in Q_j) = (\gamma + \xi \in Q_j) | \gamma \in Q_j \text{ and } -\infty \leq \xi \leq \infty\}$  does not create any underestimation or overestimation of MCS level, means the system can tolerate these prediction errors without degrading/enhancing the throughput or BER. This condition would be satisfied for all  $\xi$  which keep the predicted SNR  $\hat{\gamma}$  in the same quantization level  $Q_j$  as the actual SNR  $\gamma$ . As the feedback prediction process is completely independent of the SNR data, we define the probability of this condition as

$$P(Q_{j^0}) = P(Q_j | \gamma \in Q_j \text{ and } \hat{\gamma} \in Q_j)$$
  
= 
$$\int_{q_j}^{q_{j+1}} \int_{-\gamma}^{(q_{j+1}-q_j-\gamma)} p_{\gamma}(\gamma) p_{\xi}(\xi) d\xi d\gamma \qquad (3.21)$$

On the other hand, the condition  $\{(\hat{\gamma} \in Q_j) = (\gamma + \xi \in Q_j) | \gamma \in Q_i \text{ and } 0 \leq i < j \text{ and } \xi > 0\}$ creates overestimation of actual quantization level  $Q_i$ . This overestimation makes the system to send data with higher bitrate which increases the throughput but, at the same time, also increases the BER. We define the probability of this condition as

$$P(Q_{j^+}) = P(Q_i | \gamma \in Q_i \text{ and } \hat{\gamma} \in Q_j \text{ and } 0 \le i < j)$$

$$= \sum_{k=0}^{j-1} \int_{q_k}^{q_{k+1}} \int_{(q_j-q_k-\gamma)}^{(q_{j+1}-q_k-\gamma)} p_{\gamma}(\gamma) p_{\xi}(\xi) d\xi d\gamma \qquad (3.22)$$

Similar analysis can be done for underestimation case also. The condition  $\{(\hat{\gamma} \in Q_j) = (\gamma + \xi \in Q_j) | \gamma \in Q_i \text{ and } j < i \leq N-1 \text{ and } \xi < 0\}$  creates underestimation of actual quantization level  $Q_i$ . This makes the system to send data with lower bitrate which decreases the throughput and the BER at the same time. We define the probability of this condition as

$$P(Q_{j^-}) = P(Q_i | \gamma \in Q_i \text{ and } \hat{\gamma} \in Q_j \text{ and } j < i \le N-1)$$

$$= \sum_{k=j+1}^{N-1} \int_{q_k}^{q_{k+1}} \int_{(q_j-q_k-\gamma)}^{(q_{j+1}-q_k-\gamma)} p_{\gamma}(\gamma) p_{\xi}(\xi) d\xi d\gamma \qquad (3.23)$$

Based on this analysis, the probability that the predicted SNR  $\hat{\gamma}$  lies in quantization level j, i.e.,  $P(\hat{\gamma} \in Q_j)$  can be written as the sum of Eq. 3.21, Eq. 3.22 and Eq. 3.23; and defined after some manipulations as

$$P(\hat{\gamma} \in Q_j) = \int_{-\infty}^{\infty} \int_{q_j - \xi}^{q_{j+1} - \xi} p_{\xi}(\xi) p_{\gamma}(\gamma) d\gamma d\xi$$
(3.24)

At this step, the performance parameters can be determined using the above probability of prediction error. For discrete rate adaptation, the average spectral efficiency with prediction is given by

$$\frac{\hat{D}}{B} = \sum_{j=0}^{N-1} r_j \int_{-\infty}^{\infty} \int_{q_j-\xi}^{q_{j+1}-\xi} p_{\xi}(\xi) p_{\gamma}(\gamma) d\gamma d\xi$$
(3.25)

The average throughput with prediction is given by

$$\hat{\eta} = \sum_{j=0}^{N-1} r_j \int_{-\infty}^{\infty} \int_{q_j-\xi}^{q_{j+1}-\xi} (1 - \text{BER}(\gamma))^L p_{\xi}(\xi) p_{\gamma}(\gamma) d\gamma d\xi$$
(3.26)

The average bit error rate with prediction  $\overline{\text{BER}}$  is computed as

$$\widehat{\operatorname{BER}} = \frac{\sum_{j=0}^{N-1} r_j \int_{-\infty}^{\infty} \int_{q_j-\xi}^{q_{j+1}-\xi} \operatorname{BER}(\gamma) p_{\xi}(\xi) p_{\gamma}(\gamma) d\gamma d\xi}{\sum_{j=0}^{N-1} r_j \int_{-\infty}^{\infty} \int_{q_j-\xi}^{q_{j+1}-\xi} p_{\xi}(\xi) p_{\gamma}(\gamma) d\gamma d\xi}$$
(3.27)

Table 3.1: Minimum receiver SNR required to use in MCS

| Modulation | Level | Coding | Minimum<br>Receiver<br>SNR (db) |
|------------|-------|--------|---------------------------------|
| BPSK       | 1     | 1/2    | 3.0                             |
| OPSK       | 2     | 1/2    | 5.5                             |
|            | 3     | 3/4    | 8.0                             |
| 160AM      | 4     | 1/2    | 10.0                            |
| IUGAM      | 5     | 3/4    | 13.5                            |
| 640AM      | 6     | 2/3    | 17.0                            |
| 04QAM      | 7     | 3/4    | 18.5                            |

We investigate the performance evaluation of the downlink system of above analytical model using Matlab and Maple simulations. For error distribution with kernel density estimation (Fig. 3.7), we found the best fit with s = 0.05 and H = 64. We used these values for all the analysis of average spectral efficiency, average bit error rate and average downlink throughput.



Figure 3.8: Numerical results for actual and predicted CQI.

To calculate the downlink throughput, the MAC frame size is fixed as d = 2000 bits for all the users. The minimum receiver SNR required to decide each level of MCS according to the WiMAX standard [89] is shown in Table 3.1. The error is independent of the SNR data and from Fig. 3.7, it is observed that the error distribution is almost symmetric. These two information lead to the fact that any jump because of error from an MCS levels *i* to MCS level *j* has almost equal probability of jumping in reverse. Because of this phenomenon, we observe from Fig. 3.8a that the average spectral efficiency for actual and predicted SNR is almost same. The spectral efficiency for lower SNR shows small differences between them but the average spectral efficiency tends to be equal with higher SNR values.

From Fig. 3.8b, we observe that the average BER for both actual and predicted SNR case are same. Even though the average BERs are identical, for some situations the system BER performance could be worst because of predicted SNR. According to the Eq. 3.23, while the actual MCS level should be higher, the prediction error enables the BS to choose lower MCS level. This situation would reduce the system throughput for choosing lower MCS level but increases the reliability by decreasing the BER as well. Similarly, according to the Eq. 3.22, the prediction error enables the BS to choose higher MCS level which would increase the system throughput but decreases the reliability by increasing the BER. Fig. 3.8c shows the average downlink throughput (packet/slot/user) for actual and predicted feedback case. For lower SNR (5 dB), the average throughput with predicted feedbacks is 6.35% less than the average throughput with actual feedbacks. We observe that the average throughput with predicted feedbacks is only 0.67% less than the average throughput with actual feedbacks.

# 3.4 Opportunistic Periodic Feedback

When employing either periodic feedback strategy (i.e. IBF or PBF) and scheduling data with estimated or predicted CQI, it is merely impossible in realtime to become certain about any CQI estimation or prediction error during a uncertain period (recall fig. 3.3). If data is scheduled with over estimated CQI (i.e. higher MCS level), the BER performance is deteriorated significantly, which in turn, makes the MS unable to decode the packet correctly. In this situation, a system with ARQ or hybrid ARQ (HARQ) issues an explicit or implicit nack to the BS to retransmit the packet. While the probability of receiving an ack/nack largely depends on the channel condition, we opportunistically use this ack/nack information to reduce the effect of over or under estimations and to compensate for the unavailability of real CQI information during the uncertain period.

For this purpose, we introduce a cross-layer based opportunistic periodic feedback algorithm. We act during the uncertain period, thanks to ARQ protocol the BS does not need to wait for next CQI. In IBF or PBF, if we can detect the channel estimation as early as possible, it is possible to reduce the over or under estimation effect in frames during rest of the uncertain period. We define a  $w \times 1$  matrix  $\Psi = [\psi(t), \psi(t+1), \psi(t+2), ..., \psi(t+w-1)])$  where each element  $\psi$  is a scaling coefficient determined based on the received ack/nack. At any given frame t, we define  $\psi(t)$  as

$$\psi(t) \triangleq \begin{cases} u, & \text{if an ack is received in frame } t, \quad u \ge 1\\ v, & \text{if a nack is received in frame } t, \quad v \le 1 \end{cases}$$
(3.28)

where u and v are system parameters to determine the scaling coefficient. Parameter u is used to scale up the underestimated CQI and v is used to scale down the overestimated CQI. These scaling parameters will help in adjusting the previous memorized estimated or predicted CQI values (during the uncertain period) which are used to determine the next CQI. This opportunistic strategy helps in reducing the estimation/prediction error probability of next CQI.

### 3.4.1 Opportunistic Interval Based Feedback (OIBF)

When a CQI feedback  $\gamma(t)$  received by BS at frame t, the uncertain period lasts for next w - 1 frames when the estimated CQI could be over or under estimated. We define a  $w \times 1$  matrix  $\Gamma_{ibf} = [\gamma(t), \hat{\gamma}(t+1), ..., \hat{\gamma}(t+w-1)]$  which contains the real and estimated CQI data. In OIBF, the last w - 1 elements of  $\Gamma_{ibf}$  are equal to the first element which means  $\hat{\gamma}(t+i) = \gamma(t)$  for  $1 \leq i \leq w-1$ , as the BS uses last received CQI to decide the MCS level for next w-1 frames. From this, we define the  $w \times 1$  matrix  $E_{ibf} = [\gamma(t) - \gamma(t), \gamma(t+1) - \gamma(t), ..., \gamma(t+w-1) - \gamma(t)] = [0, e_{ibf}(t+1), ..., e_{ibf}(t+w-1)]$  which contains the CQI estimation errors calculated according to Eq. 3.5. We apply  $\Psi$  to reduce the effect of the estimation error as below

$$\Gamma_{oibf} \triangleq \Gamma_{ibf}.\Psi \tag{3.29}$$

where  $\Gamma_{oibf}$  contains the element-by-element scalar product of  $\Gamma_{ibf}$  and  $\Psi$ . As an example, if we scale  $\Gamma_{ibf}$  after receiving an ack/nack at frame t+i for  $1 \leq i < w-1$ , there are possibilities of smaller estimation error in frames t+i+1 to t+w-1.

### 3.4.2 Opportunistic Prediction Based Feedback (OPBF)

We define a  $w \times 1$  matrix  $\Gamma_{pbf} = [\gamma(t), \hat{\gamma}(t+1), ..., \hat{\gamma}(t+w-1)]$  which contains the real and predicted CQI data. The last w-1 elements of  $\Gamma_{pbf}$  are calculated according to Eq. 3.8. From this, we define the  $w \times 1$  matrix  $E_{pbf} = [\gamma(t) - \gamma(t), \gamma(t+1) - \hat{\gamma}(t+1), ..., \gamma(t+w-1) - \hat{\gamma}(t+w-1)] = [0, e_{pbf}(t+1), ..., e_{pbf}(t+w-1)]$  which contains the CQI estimation errors calculated according to Eq. 3.7. We apply  $\Psi$  to reduce the effect of the prediction error as below

$$\Gamma_{opbf} \triangleq \Gamma_{pbf}.\Psi \tag{3.30}$$

where  $\Gamma_{opbf}$  includes the element-by-element scalar product of  $\Gamma_{pbf}$  and  $\Psi$ . As an example, if we scale  $\Gamma_{pbf}$  after receiving an ack/nack at frame t+i for  $1 \leq i < w-1$ , there are possibilities of smaller prediction error in frames t+i+1 to t+w-1. The Eq. 3.8 can be written to achieve OPBF as

$$\hat{\gamma}_{opbf}(t) \triangleq \mathbf{Z}(t-1)\Gamma_{opbf}(t), \quad t \ge w$$
(3.31)

If u = 1 and v = 1 in Eq. 3.31, implies no scaling coefficient is used, hence Eq. 3.8 becomes a special case of Eq. 3.31. The OPBF algorithm is summarized in Algorithm 2.

### Algorithm 2 Opportunistic Prediction Based Feedback

1: procedure OPBF 2: input:  $\gamma(t), \quad \forall t \in \mathbb{N}$ 3: **output:**  $\hat{\gamma}_{opbf}(t), \quad t \geq w \text{ and } \forall t \in \mathbb{N}$ 4: for each scheduling time, t = w, w + 1, ... compute do  $\hat{\gamma}_{opbf}(t) = \mathbf{Z}(t-1)\Gamma_{opbf}(t)$ 5: $e(t) = \gamma(t) - \hat{\gamma}_{opbf}(t)$ 6:  $\alpha = \Gamma_{opbf}^T(t) \mathbf{M}^{-1}(t-1)$ 7:  $\beta = \lambda + \alpha \Gamma_{opbf}(t)$ 8:  $G(t) = \frac{\alpha^T}{\beta}$ 9:  $\mathbf{M}' = \mathbf{G}(t)\alpha$ 10:  $M^{-1}(t) = \frac{1}{\lambda} [M^{-1}(t-1) - M']$ 11: Z(t) = Z(t-1) + G(t)e(t)12:13: end for

3.5 Simulation Results

### 3.5.1 Simulation Settings

In this section, we evaluate the performance of our proposed opportunistic periodic feedback mechanisms via extensive simulations and gain further insight about how the parameters influence the performance. The simulations were performed using Matlab[83]. In the simulation model, system level parameters of downlink OFDMA environment are set to be compatible

| Parameters   | Value  |  |
|--|--|--|
| Channel bandwidth  | 10 MHz   |  |
| Frame duration   | $5 \mathrm{ms}$                                    |  |
| Downlink subchannelization method                        | $PUSC (1 \text{ channel} \times 2 \text{ symbol})$ |  |
| Uplink subchannelization method                          | PUSC (1 channel $\times$ 3 symbol)                 |  |
| # of downlink subchannels                                | 30   |  |
| # of uplink subchannels                                  | 35   |  |
| Downlink:Uplink Ratio                                    | 2:1  |  |
| # of downlink symbols                                    | 27   |  |
| # of uplink symbols                                      | 18   |  |
| Slow Mobility Channel                                    | ITU Ped-A 3 km/h                                   |  |
| Fast Mobility Channel                                    | ITU Veh-B 60 km/h                                  |  |
| Cell radius  | 1 km   |  |
| Path loss PL (d)   | $12\log(4\pi d/\lambda) - 27$                      |  |
| Maximum number of retransmission in ARQ $(N_{re}^{max})$ | 4  |  |
| VoIP packet size per frame                               | 12 bytes   |  |
| Video packet size per frame                              | 121 bytes  |  |

Table 3.2: Simulation parameters

with the IEEE 802.16e standard [89]. We consider a single-cell, single-sector system with 10 MHz bandwidth where BS communicates with 30 MSs. In each scenario, MSs are randomly distributed in the cell with cell radius 1 km. For all the simulated scenarios, we conduct a set of experiments using two channel patterns indicative of the typical Mobile WiMAX mobility profiles, the ITU pedestrian A model representing low mobility scenario of 3 km/h, and the ITU vehicular B model corresponding to high mobility scenario of 60 km/h.

Each MS has a data session of 156 seconds equivalent to 31200 frames. During a simulation scenario, 50% of the total users are video users and rest 50% are VoIP users. Video and VoIP users have a data rate of 176 kbps with H.264 format and 5.3 kbps with G723.1 [101] respectively. Considering MAC header and fragmentation or packing header overhead, the data rates are 189 kbps (121 bytes per frame) and 18.75 kbps (12 bytes per frame) for video and VoIP users respectively. At this step, we use a simple downlink scheduling mechanism at the BS as in [101] where each MS is scheduled in each frame. More advanced downlink scheduling mechanisms will be used in next chapters. Each user k sends its feedback periodically in uplink based on its feedback window  $w_k$ . In case of a packet loss, the ARQ protocol uses maximum number of retransmission  $N_{re}^{max} = 4$  to send the packet, otherwise the packet is considered as lost. We run the simulation 30 times with different seeds then we calculate the means and 95% confidence intervals to plot the curves. The minimum received SNR required to decide each MCS level shown in Table 3.1 is obtained by simulations setting target BER =  $10^{-3}$ .



Figure 3.9: Average BER for different feedback windows.



Figure 3.10: Average BER of under estimated cases for different feedback windows.



Figure 3.11: Average BER of over estimated cases for different feedback windows.

### 3.5.2 Performance Evaluation

We want to show the effect of scaling parameters u and v and the performance of our opportunistic scheme. We start with a scenario consisting 30 MSs. As we have seen in Figure 3.5 that the CQI prediction is better than using last received CQI, PBF method generates smaller average BER compared to IBF shown in Figure 3.9.

To get a more in-depth insight on the property of estimation errors, we study the BER of under and over estimation cases separately. In Figure 3.10, we analyze all under estimated cases and compare the average BER of IBF and PBF methods with actual BER. Similar analysis has been undertaken for overestimation cases as shown in Figure 3.11. As slow mobility users experience smaller error variance, the difference between the actual BER and IBF/PBF BER is less significant compared to fast mobility users scenarios. Among the underestimated cases shown in Figure 3.10, average BER is not really affected by feedback window variation which is true for both mobility cases. This is due to the fact that the variation of BER values below  $10^{-4}$  is not noticeable. On the other hand, among the overestimated cases shown in Figure 3.11, the high mobility users experience increasing average BER with the increase of feedback windows, while average BER is almost constant for different feedback windows in slow mobility cases. We also observe that the average BER is higher than target BER  $10^{-3}$  which acts on higher packet loss rate.



Figure 3.12: Average throughput results for different values of u (% of increase) and v (% of decrease).



Figure 3.13: MAC frame loss rate for different values of u (% of increase) and v (% of decrease).

To examine the effect of our scaling coefficient parameters, we show the performance results of OIBF and OPBF with varying u and v simultaneously. Note that the reference IBF and PBF scheme corresponds to the case with 0% scaling where no opportunistic behavior is considered. Figure 3.12, Figure 3.13, and Figure 3.14 present the average throughput, MAC layer packet



Figure 3.14: MAC frame retransmission rate for different values of u (% of increase) and v (% of decrease).

loss rate, and retransmission rate, respectively for varying u and v where u increases and v decreases identically. All the figures shows the results for pedestrian and vehicular mobilities. From the results, it is trivial to show that the opportunistic mechanism is not useful for low mobility case. As the channel variation is low in pedestrian mobility, the application of scaling parameters further deteriorates the channel rather than improving.

On the other hand, our opportunistic mechanism shows performance improvement for vehicular mobility case. The average throughput (Figure 3.12b) achieves the global maximum at 5% scaling. We can also observe that this scaling of combined u and v corresponds to the optimal MAC frame lossrate (Figure 3.13b) and retransmission rate (Figure 3.14b) for the considered scenario. The reduction in retransmission rate has a positive effect on maintaining the QoS of delay-sensitive applications. It is worth observing that the performance improvements is bigger for higher feedback window w = 8 compared to lower feedback windows w = 4 or w = 2.

### 3.5.3 Space and Memory Complexity for Prediction Algorithm

The practical implementation of our PBF scheme requires the integration of the prediction module at the BS side. For each active MS, the BS has to handle a dedicated prediction module. The complexity grows linearly with the number of active MS. The space complexity depends on the prediction window for each MS. Each instance of the algorithm occupies  $4 + 4w + 4w^2$  unit double type or  $O(w^2)$  space. Apart from the prediction window w, the time complexity depends on the matrix and vector multiplication. In Algorithm 1, we showed the time complexity line by line. We used flops as an abbreviation for one addition (subtraction) + one multiplication (floating point operation). The overall complexity of the algorithm is  $O(w^2)$  operations (flops) per time iteration. If we assume that the largest prediction window w = 8 as in Mobile WiMAX [89] [90], for each MS, the space complexity of our algorithm is upper bounded by  $O(8^2)$  and time complexity  $O(8^2)$ .

# 3.6 Conclusion

In this chapter, we propose a novel mechanism to reduce the feedback overhead in multiuser multicarrier OFDMA systems. We propose prediction based feedback (PBF) which allows the BS to predict the CQI feedbacks based on RLS algorithm. The PBF reduces the error variance compared to interval based feedback (IBF) and reduces over and under estimation variations. We show that it is useful to use channel prediction as a tool to reduce the feedback overhead.

Then we introduce the analytical model for system performance with and without prediction. We define the error distribution from the error data to derive the probability of error in each adaptive modulation level and some performance metrics with both predicted CQI and actual CQI. Our analytic model and numerical results shows the throughput degradation in downlink channel created by the lack of CQI feedback. The performance degradation becomes negligible for channels with high SNR.

In order to mitigate the possible under and over estimation effects of CQI prediction, we propose an opportunistic periodic feedback mechanism. In this mechanism, we exploit the hints from MAC-layer ARQ protocol to enhance the performance of periodic feedback mechanisms. The opportunistic mechanism improves the system performance for high mobility cases compared to low mobility cases.

However, this chapter deals with fixed feedback window for each MS. As the channel condition for the MSs changes over time, having a fixed feedback window is not a optimal solution. In current systems, the operators reserve a fixed number of slots for CQI feedback and it does not change frame-by-frame. If the number of users is greater than number of CQI slots, users compete for the CQI resources. In order to address these issues, we present a novel and complete framework in the next chapter.

# Chapter 4

# Adaptive Feedback Resource Allocation and Prediction

# 4.1 Introduction

In this chapter, we deal with the CQI allocation problem in a multi-user environment with strict feedback budget constraint. Our main objective is to determine the users priority to send CQI feedback in a heterogeneous environment and distribute the CQI resource to the users in each scheduling frame. To achieve this goal, we propose an integrated cross-layer framework named FEREP (Feedback Resource Allocation and Prediction). The proposed framework, implemented at the BS side, is composed of three modules: the feedback window adaptation (FWA), the priority-based feedback scheduling (PBFS) and the prediction based feedback (PBF). In the FWA module, the feedback window size  $w_k$  is tuned based on the received ack/nack from Automatic Repeat Request (ARQ) protocol that implicitly reflects the current channel condition. The PBFS module then performs feedback scheduling by taking into account the feedback window size, the user profile (heterogeneous among MSs), and the total system feedback budget. The PBF module performs CQI prediction by using recursive least square (RLS) algorithm when the channel feedback is not scheduled in current frame to choose proper MCS level.

Our contribution of this work is threefold. First, we propose a novel and practical framework for feedback resource allocation in OFDMA system with total CQI feedback budget constraint. Second, our framework takes into account, for the first time, to the best of our knowledge, service differentiation in the CQI feedback allocation strategy. Finally, our framework enhances the system performance by exploiting the cross-layer strategies and the CQI prediction tool in the absence of real CQI feedback.

## 4.2 Related Works

The challenge of reducing feedback overhead started mostly with the evolution of multiuser systems to exploit the multiuser diversity [72] [46]. The overhead increased exponentially in multiuser multicarrier systems like OFDMA [25] [52] [75] [63] and with the use of advanced technologies like MIMO [109] [100] [69] [95] [81].

An important line of work for feedback reduction in multiuser diversity is opportunistic feedback [46] [54] [47] [26] [52] [38] [95] [4] [3] [125]. Each user having CQI value above some known threshold value are allowed to send feedback. It has been proved very efficient in the case of serving a few number of MSs out of a large number of simultaneously active users. The opportunistic method is applied to a multichannel system in [26] [52] [95] [125] to have feedbacks for best-n subcarriers. This scheme is designed for frequency-selective fading channels where contiguous subcarriers are assigned to each MS and mainly used for low mobility cases [20]. It also carries huge overhead as every MS needs to send CQI for n subcarriers. The schemes [89] [86] [76] [64] designed for distributed subcarriers allow to send one averaged CQI per MS and generate less feedback overhead compared to previous schemes.

In a multicarrier multiuser system, the opportunistic feedback scheme will have no effect in feedback reduction if all the MSs need to be scheduled in each scheduling slot. An OFDMA system (e.g. WiMAX) with 10 MHz frequency bandwidth and simple scheduling can support upto 82 voice-over-IP (VoIP) users in each scheduling slot [101]. Scheduling the feedback of these MSs with 2:1 DL and UL frame ratio consumes 40% of uplink bandwidth [45], besides other overheads from ARQ ack and periodic ranging. To overcome this problem, [89] employs interval-based feedback (IBF) [89] [112] [31] which is based on a feedback window w. Each MS sends feedback periodically in round robin fashion once in every w frames and uses this *received feedback* in next (w - 1) frames as *last received feedback*. Hyeju Oh et al. [86] proposed an adaptive CQI feedback period algorithm based on user mobility in terms of doppler frequency. The feedback window for each MS is estimated according to the doppler frequency experienced by the MS. Fast users are assigned smaller feedback windows and vice-versa. Iijima et al. [64] extended the work in [86]. They adapt the feedback period to different time slots according to different MCS while keeping the packet size and scheduling time fixed. Both schemes have to synchronize based on maximum doppler frequency experienced in the system. Allocating the scheduling and feedback period based on doppler spread may not be a good idea. Doppler effect cannot track the MSs which are moving perpendicular to the BS. Besides, it is possible that the MSs experience good channel condition while moving in faster speed, or vice-versa.

Prediction based feedback (PBF) is proposed in [14]. Unlike the IBF, the PBF predicts the feedback for next (w - 1) frames instead of using last received feedback. [14] showed that it is possible to reduce the feedback load which in terms increases the uplink capacity. However, the lack of feedback affects also the downlink capacity. [13] analyzed and showed that downlink degradation is almost negligible compared to the uplink gains. In those works, we did not take into account any hints from MAC-layer ARQ protocol. Allocating all the uplink capacity for CQI feedback is too inefficient from a network operators point of view [23] [45]. For a better resource optimization, the operators gain in limiting CQI overhead to some percentage of the total uplink capacity or the total number of MSs [124] [73].

Compared with existing work, we develop a cross-layer framework of adaptive feedback with prediction for OFDMA systems. Our contribution lies in the integration of the feedback window adaptation and channel condition prediction in the resource-limited environment with total feedback budget constraint, which shows significant performance gain as demonstrated by simulation performed in both scenarios, especially in the case with stringent feedback budget constraint. In our work, we assumed the perfect CQI reception by at the BS. The studies with imperfect CQI are done in [104] [63] [111] [91]. The emission of CQI channel has been proposed in [32] [7] by exploiting ARQ protocol. They used the signalization of ARQ protocol for link adaptation. In this work, we used cross-layer information from ARQ protocol to adapt the feedback window for each user [15] [16].

The channel correlation was proposed in [34] [33] to reduce the feedback overhead. They proposed long-range prediction of subcarrier correlations so that one feedback can represent the correlated feedbacks. The prediction also used in [67] to predict MIMO beamforming. The channel prediction in frequency domain has been studied in [11]. Channel prediction in time domain has been studied in [13] [36] [97] [65] [35] [99] [21]. We proposed CQI prediction using RLS algorithms as in [13].



Figure 4.1: OFDMA system

# 4.3 System Model

The system has a set  $\mathcal{K}$  of K simultaneously active mobile stations (MSs, also referred to as users throughout the chapter) communicating with a single base station (BS) as shown in Figure 4.1. To allow the BS to apply adapted MCS, MSs periodically send channel information to the BS. The BS maintains a set  $\mathcal{W}$  of feedback window  $w_k$  for each MS k and allows MS kto send channel feedback once in each  $w_k$  frames. The system also has a total feedback budget constraint, denoted as F, meaning that each frame can carry the feedback of at most F MSs defined as set  $\mathcal{F}$ . In other words, at most F MSs can send feedback in a frame. If the number of MSs scheduled to send their feedback is larger than F, the BS should admit at most Famong them to send feedback.

Modern multiuser multicarrier OFDMA systems are aimed to support user heterogeneity. In our model, each MS k is characterized by its service priority, denoted by  $\alpha_k$ , an operator defined parameter which represents the aggregation of several users priority metrics such as QoS requirements, user profile, and the amount of payment to the operator, etc. For Mobile WiMAX systems, service differentiation is one of the key elements. Operators are supposed to propose diversified offers for subscribers while maximizing system performance, e.g., subscribers can be virtually grouped by the operator according to different metrics (e.g. high/low tariffs, private/corporate subscriptions, sensibility/tolerance to QoS degradations) and provide for each group dedicated and differentiated treatments. In this regard, MSs with higher  $\alpha$  values are supposed to get better service than those with lower  $\alpha$  values.

# 4.4 Feedback Resource Allocation and Prediction (FEREP)

As mentioned previously, in current standardized OFDMA systems [89] [2] [108], each MS is assigned a feedback window without taking into account user heterogeneity in terms of application, mobility, QoS constraint, etc. However, today's wireless systems are typically composed of MSs with different applications (e.g. VoIP and video) and mobility profiles (e.g. vehicular and pedestrian). In such context, applying a homogeneous feedback mechanism to heterogeneous MSs is clearly non-adaptive. The situation deteriorates in the case where there is a stringent constraint on the total number of feedbacks sent by MSs due to the limited radio resource in OFDMA systems (i.e., only F CQI slots are dedicated for K MSs to send feedback while usually K > F). This motivates our proposition of a cross-layer framework of adaptive feedback with prediction, consisting of three modules: the feedback window adaptation (FWA), the priority-based feedback scheduling (PBFS) and the prediction based feedback (PBF).

At the beginning of each frame, FWA algorithm at the BS dynamically calculates the feedback window for each MS based on the ARQ signaling with explicit ack/nack, received for the packets transmitted in previous frames. As there are only F CQI slots in each frame available to send feedback, the PBFS algorithms selects the F out of K MSs to send feedback in that frame based on a weighted-priority algorithm. The rest of the (K - F) MSs are not scheduled to send feedback in that frame, instead, their CQIs are predicted by the PBF module based on recursive least square (RLS) algorithm. Figure 4.2 gives a system-level overview of the proposed framework and represents the interaction between physical and MAC layers. In the following, we provide a detailed analysis on the proposed framework modules.

### 4.4.1 Feedback Window Adaptation (FWA)

The objective of the FWA algorithm is to dynamically calculate the appropriate feedback window size based on MSs' channel condition. As explained in related works, using the doppler frequency as metric [64] [86] may be inappropriate as a MS experiencing high doppler frequency may still have sufficiently good channel to recover lost packets correctly. Another point is, in existing systems where CQI is sent once in every w frames (w > 1), the BS has no knowledge about the channel change within the interval between two successive CQIs. In FWA, we exploit the ack/nack packets of the ARQ protocol which are returned in every frame to adjust the feedback window size.

More specifically, we assume that the ARQ protocol employs go-back-n mechanism. If a



Figure 4.2: FEREP Framework

packet is recovered by the MS properly, it sends an ack frame back to the BS. Otherwise, the MS sends a nack frame and the BS retransmits the packet<sup>1</sup>. The ARQ protocol maintains a parameter for maximum number of retransmissions  $N_{re}^{max}$ . When a packet is retransmitted  $N_{re}^{max}$  times and no ack is received, it is considered lost.  $N_{re}^{max}$  is application dependant and increasing its value would violate the delay constraint for an application [80].

The proposed FWA algorithm based on the above ARQ model implemented at the base station to calculate the feedback window size for MS k (denoted as  $w_k$ ) is shown in Algorithm 3. The core idea is to increase the feedback window size of a MS when its channel condition is good, reflected by consecutively received ACKs, and decrease the feedback window size once a nack is received, implying a possible deterioration of the channel quality. Note that a small feedback window size means that the real measured CQI is returned to BS more frequently (maximum, is sending in every frame) and larger window size relies on more predicted CQI at the expense of a higher risk of alteration due to the prediction error (which effect is more negligible in good channel condition [13]).

To this end, initially (line 1),  $w_k$  is initialized to  $w_k^{init}$ . Upon receiving an ack from MS k, the BS increases the counter  $n_{ack}$  which memorizes the number of consecutive received ACKs (lines 3 – 4). If the consecutive number of received ack exceeds the threshold  $N_{ack}^{th}(\alpha_k)$ 

 $<sup>^{1}</sup>$ Note that our mechanism works also with implicit ARQ. In this case, retransmissions are performed after time-out.

### 4.4. FEEDBACK RESOURCE ALLOCATION AND PREDICTION (FEREP47

**Algorithm 3** Feedback Window Adaptation procedure to calculate  $w_k$ 

1: initialization: Set  $N_{ack}^{th}(\alpha_k), w_k \leftarrow w_k^{init}, n_{ack} \leftarrow 0$ , 2: loop if ack received then 3:  $n_{ack} \leftarrow n_{ack} + 1$ 4: if  $n_{ack} \geq N_{ack}^{th}(\alpha_k)$  then 5: $w_k \leftarrow w_k + 1$ 6:  $n_{ack} \leftarrow 0$ 7: end if 8: 9: else if nack received or timeout then  $w_k \leftarrow \max\{w_k - 1, 1\}$ 10: $n_{ack} \leftarrow 0$ 11:end if 12:13: end loop

which depends on the service priority of MS  $k \alpha_k$ , then the BS increases  $w_k$  by 1 and  $n_{ack}$ is reset to 0 (lines 5 - 8). Otherwise upon receiving a nack, the BS reduces  $w_k$  by 1(lines 9 - 12) and  $n_{ack}$  is reset to 0. A desirable property of the proposed FWA algorithm is that by tuning the parameters  $N_{ack}^{th}(\alpha_k)$ , the BS can achieve a balance between the robustness and feedback overhead, e.g., a larger value of  $N_{ack}^{th}(\alpha_k)$  leads to a more conservative increase in  $w_k$ , the algorithm is thus more robust at the price of more feedback overhead caused by potential under-estimation of the feedback window size.

# 4.4.2 Priority-based Feedback Scheduling (PBFS) under Total Budget Constraint

Due to limited radio resource, Mobile WiMAX systems impose a budget constraint on the total number of feedbacks sent by the MSs to the BS. A critical question in this context is which MSs should be admitted to send feedback given the total budget constraint. In this subsection, we establish a priority-based feedback scheduling algorithm, in which the admissible MSs allowed to send feedback at frame t are chosen based on their priority according to their feedback window size, the last feedback sending time and their service priority  $\alpha_k$ .

The proposed PBFS algorithm, performed after the FWA algorithm, determines which MSs

are admitted to send CQI at frame t based on the following metric  $\{Y_k, k \in \mathcal{K}\}$ :

$$Y_k(t) \triangleq g(\alpha_k)(t - t_k^{last})/w_k, \quad k \in \mathcal{K},$$
(4.1)

where  $g(\alpha_k)$  is a function of the service priority of MS k, t is the current time,  $t_k^{last}$  is the time of the last received feedback from the MS k, and  $w_k$  is feedback window size of MS k. The priority score is thus the priority  $(t - t_k^{last})/w_k$ , determined by the feedback window and the last serving time, weighted by the user-dependent service priority. The PBFS algorithm is executed at the BS in the case where the number of MSs scheduled to send their feedback exceeds the total budget F. In this case, the PBFS algorithm admits the F MSs with highest priority scores. By doing so, the BS prioritizes MSs with bad channel condition (small feedback window size) and higher service priority while maintaining certain fairness by integrating  $t_k^{last}$  in  $Y_k$ .

To conclude this subsection, it is insightful to study the proposed PBFS algorithm from the perspective of scheduling. To this end, rewrite  $Y_k$  as

$$Y_k = \frac{g(\alpha_k)}{w_k} (t - t_k^{last}).$$

$$(4.2)$$

The PBFS algorithm can be essentially viewed as the round-robin scheduling scheme weighted by coefficient  $g(\alpha_k)/w_k$  depending on the users' service priority and feedback window size. Consequently, for a MS k for which the feedback should be scheduled (i.e.,  $w_k < t - t_k^{last}$ ),  $\frac{g(\alpha_k)}{w_k}F$  $\frac{g(\alpha_k)}{w_k}F$  slots are allocated in average among the total budget F. In the degenerated case where  $g(\alpha_k)/w_k$  takes the same value for all MSs, the PBFS algorithm becomes the classical round-robin scheduling that ensures absolute fairness among MSs in terms of feedback resource allocation.

### 4.4.3 Prediction Based Feedback (PBF)

The PBF algorithm predicts the channel condition, i.e., SNR, when the feedback is not scheduled in the current frame t. To this end, we integrate the prediction mechanism we proposed in chapter 3. The algorithm is shown again in Algorithm 4.

### 4.5 Performance Metric for Feedback Window Adaptation

In this section, we derive the performance metric of our FWA algorithm. We intend to derive probability of receiving nack P(nack) which acts on increasing the feedback window.

Algorithm 4 Opportunistic Prediction Based Feedback

1: procedure OPBF

2: **input:**  $\gamma(t)$ ,  $\forall t \in \mathbb{N}$ 

3: **output:**  $\hat{\gamma}_{opbf}(t), \quad t \geq w \text{ and } \forall t \in \mathbb{N}$ 

4: for each scheduling time, t = w, w + 1, ... compute do

- 5:  $\hat{\gamma}_{opbf}(t) = \mathbf{Z}(t-1)\Gamma_{opbf}(t)$
- 6:  $e(t) = \gamma(t) \hat{\gamma}_{opbf}(t)$
- 7:  $\alpha = \Gamma_{opbf}^T(t) \mathbf{M}^{-1}(t-1)$
- 8:  $\beta = \lambda + \alpha \Gamma_{opbf}(t)$
- 9:  $G(t) = \frac{\alpha^T}{\beta}$
- 10:  $M' = G(t)\alpha$
- 11:  $M^{-1}(t) = \frac{1}{\lambda} [M^{-1}(t-1) M']$
- 12: Z(t) = Z(t-1) + G(t)e(t)

13: end for

From P(nack) we can derive P(ack), the probability of receiving an ack which is used to increase the feedback window.

### 4.5.1 Interval Based Feedback (IBF) Case

In IBF, BS uses the most simplest mechanism to employ partial feedback. BS uses the last received CQI to decide the MCS level for next  $w_k - 1$  frames, where  $w_k > 1, \forall k$ . Any probable change in the channel condition will enable the user to experience over or under estimation of the MCS level. In Figure 4.3a, we can see that actual channel conditions at times t and  $t + T_f$ fall in two MCS quantization levels  $Q_j$  and  $Q_{j-1}$  respectively. When CQI  $\gamma_k(t)$  of user k is received at time t, IBF creates overestimation at time  $t + T_f$  when user k is scheduled using last MCS level  $Q_j$  instead of actual MCS level  $Q_{j-1}$ . The probable SNR estimation error can be calculated from the difference between last received SNR at time t and the current channel condition at time  $t + dT_f$ . As all the following analysis holds for all MSs, we drop the subscript k.

If the SNR  $\gamma$  at time t is known, and we expect to estimate SNR  $\hat{\gamma}$  in next frame after time  $dT_f$ , the distribution of  $\hat{\gamma}$  conditioned on  $\gamma$  is in turn a non-central Chi-squared (NC $\chi^2$ ) distributed random variable with two degrees of freedom having pdf [92]

$$f_{\gamma}(\hat{\gamma}|\gamma) = \frac{\sigma_n^2}{\sigma_e^2} e^{-\frac{\sigma_n^2}{\sigma_e^2}(\hat{\gamma}+\rho^2\gamma)} I_0(2\frac{\sigma_n^2}{\sigma_e^2}\sqrt{\hat{\gamma}\rho^2\gamma})$$
(4.3)

where  $I_0$  is the *zero*th order modified Bessel function of the first kind,  $\sigma_n^2$  gaussian noise variance, and  $\sigma_e^2$  is the prediction error variance. Wong et al. [63] approximated the above NC $\chi^2$  distribution using Gamma distribution as

$$f_{\gamma}(\hat{\gamma}|\gamma) \approx \frac{\beta^{\alpha}}{\Gamma(\alpha)} \hat{\gamma}^{\alpha-1} e^{-\beta\hat{\gamma}}$$
(4.4)

where  $\alpha = (\kappa \rho^2 + 1)^2 / (2\kappa \rho^2 + 1)$  is the Gamma pdf shape parameter with  $\kappa = \gamma \frac{\sigma_n^2}{\sigma_e^2}$ , and  $\beta = \alpha / (\gamma \rho^2 + \frac{\sigma_e^2}{\sigma_e^2})$  as scale parameter.

Let define the probability of experiencing  $\hat{\gamma}(t + dT_f) \in Q_{j-1}$  conditioned on  $\gamma(t) \in Q_j$  as  $P(\gamma(t + dT_f) \in Q_{j-1} | \gamma(t) \in Q_j)$ . To derive this probability, we first need to calculate the probability of  $\hat{\gamma}(t + dT_f) \in Q_{j-1}$  conditioned on last received CQI  $\gamma(t)$  regardless any MCS level. Let define this probability as  $P(\hat{\gamma}(t + dT_f) \in Q_{j-1} | \gamma(t))$ . Using [50, Section 3.383.3], one can calculate the probability of  $\hat{\gamma}(t + dT_f)$  falling above some SNR threshold  $q_j$  as

$$P(\hat{\gamma}(t+dT_f) \ge q_j | \gamma(t)) = \int_{q_j}^{\infty} \frac{\beta^{\alpha}}{\Gamma(\alpha)} \hat{\gamma}^{\alpha-1} e^{-\beta\hat{\gamma}} d\hat{\gamma}$$
$$= \frac{\Gamma(\alpha, q_j / \beta)}{\Gamma(\alpha)}$$
(4.5)

where  $\Gamma(.,.)$  is the incomplete Gamma function, and  $\Gamma(.)$  is the Gamma function. Hence,  $P(\hat{\gamma}(t + dT_f) \in Q_{j-1}|\gamma(t))$  is calculated as

$$P(\hat{\gamma}(t + dT_f) \in Q_{j-1} | \gamma(t))$$

$$= P(\hat{\gamma}(t + dT_f) \ge q_{j-1} | \gamma(t)) - P(\hat{\gamma}(t + dT_f) \ge q_j | \gamma(t))$$

$$= \frac{\Gamma(\alpha, q_{j-1} / \beta)}{\Gamma(\alpha)} - \frac{\Gamma(\alpha, q_j / \beta)}{\Gamma(\alpha)}$$
(4.6)

According to Equation 3.16, the BER increases exponentially with the size of modulation constellation for a fixed SNR. Consequently, the PER also increases exponentially and is largely responsible for not recovering the MAC frame properly by MAC layer, hence produces nack.

Now we are ready to derive the probability of receiving a nack or ack. For simplicity, we assume that the system always maintains the target BER and no nack is generated for w = 1. A nack may be generated when the MCS is overestimated. Once we know the probability of having a *nack*, we can tackle the overestimation by reducing the w to increase the rate of CQI



Figure 4.3: Overestimation in (a) interval based feedback (IBF) (b) prediction based feedback (PBF)

feedback. Using Equation 4.6, we can derive the  $P_{ibf}(nack)$ , the probability of receiving a nack in IBF scheme as

$$P_{ibf}(nack) = \frac{1}{m} \sum_{j=1}^{m} P(\hat{\gamma}(t + dT_f) \in Q_{j-1} | \gamma(t) \in Q_j) \\ = \frac{1}{m} \sum_{j=1}^{m} \int_{q_j}^{q_j+1} P(\hat{\gamma}(t + dT_f) \in Q_{j-1} | \gamma(t)) d\gamma$$
(4.7)

with  $q_1 = 0, q_{m+1} = \infty$  and we derive  $P_{ibf}(ack)$ , the probability of receiving an ack in IBF

scheme as

$$P_{ibf}(ack) = 1 - P_{ibf}(nack) \tag{4.8}$$

The feedback window is increased by one after  $N_{ack}^{th}$  number of consecutive acks are received with probability of  $\prod_{i=1}^{N_{ack}^{th}} P_{ibf}(ack)$  if we consider the errors are independent.

### 4.5.2 Prediction Based Feedback (PBF) Case

In prediction based feedback (PBF), the BS uses a linear predictor based on RLS algorithm to predict the CQI. Unlike IBF method, this predicted SNR is used instead of last received SNR to schedule data in next w - 1 frames, where w > 1. In Figure 4.3b, we can see the real and predicted channel conditions at times  $t + T_f$  fall in two MCS quantization levels  $Q_j$  and  $Q_{j-1}$  respectively. Once the last CQI is received at time t, the overestimation at time  $t + dT_f$ is estimated with respect to the difference between approximated real SNR  $\hat{\gamma}(t + dT_f)$  and predicted SNR  $\tilde{\gamma}(t + dT_f)$ . We can estimate  $\hat{\gamma}(t + dT_f)$  using the distribution in Equation 4.3. Now we have to estimate the predicted SNR  $\tilde{\gamma}(t + dT_f)$  conditioned on  $\hat{\gamma}(t + dT_f)$ . Let us define the prediction error of SNR predictor as  $\xi$ . As there is no correlation between the channel condition and prediction error [94][36], the predicted SNR is

$$\tilde{\gamma} = \hat{\gamma} + \xi \tag{4.9}$$

where  $\xi$  has a Gaussian distribution with variance  $\sigma_{\xi}^2$ .

To determine the overestimation, we determine  $P(\tilde{\gamma}(t + dT_f) \in Q_j | \hat{\gamma}(t + dT_f) \in Q_{j-1})$ , the probability of  $\tilde{\gamma}(t + dT_f) \in Q_j$  conditioned on  $\hat{\gamma}(t + dT_f) \in Q_{j-1}$  as [13]

$$P(\tilde{\gamma}(t+dT_f) \in Q_j | \hat{\gamma}(t+dT_f) \in Q_{j-1}) = \int_{q_j-q_{j-1}}^{q_{j+1}-q_j} \left[ \int_{q_{j-1}}^{q_j} f_{\xi}(\xi) f_{\hat{\gamma}}(\hat{\gamma}) d\hat{\gamma} \right] d\xi$$
(4.10)

where  $f_{\xi}(\xi)$  is the probability density function of  $\xi$  and  $f_{\hat{\gamma}}(\hat{\gamma})$  is, in fact,  $f_{\gamma}(\hat{\gamma}|\gamma)$  from Equation 4.3. Hence, for last received SNR  $\gamma(t) \in Q_l$ , we can write

$$P(\tilde{\gamma}(t+dT_f) \in Q_j | \hat{\gamma}(t+dT_f) \in Q_{j-1} \text{ and } \gamma(t) \in Q_l)$$

$$= \int_{q_l}^{q_{l+1}} \left[ \int_{q_j-q_{j-1}}^{q_{j+1}-q_j} \left[ \int_{q_{j-1}}^{q_j} f_{\xi}(\xi) f_{\hat{\gamma}}(\hat{\gamma}|\gamma) d\hat{\gamma} \right] d\xi \right] d\gamma$$
(4.11)

Using Equation 4.11, the probability of receiving a nack in PBF scheme is  $P_{pbf}(nack)$ 

| Parameters   | Value                              |  |
|--|------------------------------------|--|
| Channel bandwidth  | 10 MHz                             |  |
| Frame duration   | $5 \mathrm{ms}$                    |  |
| Downlink subchannelization method                        | PUSC (1 channel $\times$ 2 symbol) |  |
| Uplink subchannelization method                          | PUSC (1 channel $\times$ 3 symbol) |  |
| # of downlink subchannels                                | 30                                 |  |
| # of uplink subchannels                                  | 35                                 |  |
| Downlink:Uplink Ratio                                    | 2:1                                |  |
| # of downlink symbols                                    | 27                                 |  |
| # of uplink symbols                                      | 18                                 |  |
| Slow Mobility Channel                                    | ITU Ped-A 3 km/h                   |  |
| Fast Mobility Channel                                    | ITU Veh-B 60 km/h                  |  |
| Cell radius  | 1 km                               |  |
| Path loss PL (d)   | $12\log(4\pi d/\lambda) - 27$      |  |
| Maximum number of retransmission in ARQ $(N_{re}^{max})$ | 4                                  |  |
| VoIP packet size per frame                               | 11.5 bytes                         |  |
| Video packet size per frame                              | 121 bytes                          |  |

Table 4.1: Simulation parameters

defined as

 $P_{pbf}(nack)$ 

$$= \frac{1}{m} \sum_{j=1}^{m} \mathbb{P}(\tilde{\gamma}(t + dT_f) \in Q_j | \hat{\gamma}(t + dT_f) \in Q_{j-1} \text{ and } \gamma(t) \in Q_l)$$

and the  $P_{pbf}(ack)$ , the probability of receiving an ack in PBF scheme as

$$P_{pbf}(ack) = 1 - P_{pbf}(nack) \tag{4.12}$$

The feedback window is increased by one after  $N_{ack}^{th}$  number of consecutive acks are received with probability of  $\prod_{i=1}^{N_{ack}^{th}} P_{pbf}(ack)$  if we consider the errors are independent.

# 4.6 Simulation Study
| Modulation | Level | Coding | Minimum Receiver SNR (db) | Bitrate |
|------------|-------|--------|---------------------------|---------|
| BPSK       | 1     | 1/2    | 3.0                       | 1       |
| QPSK       | 2     | 1/2    | 6.0                       | 2       |
|            | 3     | 3/4    | 8.5                       | 2       |
| 16QAM      | 4     | 1/2    | 11.5                      | 4       |
|            | 5     | 3/4    | 15.0                      | 4       |
| 64QAM      | 6     | 2/3    | 19.0                      | 6       |
|            | 7     | 3/4    | 21.0                      | 6       |

Table 4.2: Minimum receiver SNR required to use in MCS

In this section, we evaluate the performance of the proposed framework via extensive simulations using Matlab [83] and gain further insight on how different system parameters influence the performance. More specifically, we study the behavior of the proposed framework in both homogenous scenario where users have the same service priority  $\alpha$ , and heterogeneous scenario where users have different service priorities. For both scenarios, we conduct a set of simulations using two channel patterns indicative of the typical Mobile WiMAX mobility profiles, the ITU pedestrian A model representing low mobility scenario of 3km/h, and the ITU vehicular B model corresponding to high mobility scenario of 60km/h.

#### 4.6.1 Simulation Setting

We study a single-cell, single-sector system where the BS communicates with K MSs randomly distributed in the cell with radius 1km. The network parameters, as shown in Table 1, are set following the IEEE 802.16e standard [89]. As mentioned previously, two channel mobility models are simulated: ITU pedestrian A (3km/h) and ITU vehicular B (60km/h). Each MS has a session of 156 seconds equivalent to 31200 frames (200 frames/sec for 5ms frame duration). Among the MSs, 50% of them are voice over IP (VoIP) users and other 50% are video users. The VoIP users have data rate of 5.3 kbps with G723.1 Annex A format [101]. The video users have data rate of 176 kbps with H.264 format. Taking consideration of MAC header and fragmentation or packing header overhead, the data rates are 18.75 kbps and 189 kbps for VoIP and video users, respectively. We assume a simple downlink scheduling algorithm as in [101] where each MS is scheduled in each frame. The minimum received SNR required to decide each MCS level is shown in Table 2 [89]. Each simulation is run 10 times with different seeds.

#### 4.6.2 Homogenous Scenario

We start with the homogenous scenario consisting of K = 30 MSs with the same service priority  $\alpha$ . In Figure 4.4, we show the effect of feedback constraint parameter F on the performance of the proposed framework with the IBF reference scheme in terms of average goodput<sup>2</sup>. In this scenario, F out of K MSs send feedback in each frame, and every MS has same priority to send feedback. In the reference scheme, MSs are scheduled to send feedback in round robin fashion. MSs who are not scheduled to send feedback, their last received feedbacks are used as the current channel condition according to the IEEE 802.16e standard [89] or works in [86] [112] [31]. Note that when F = K, the reference scheme as well as FEREP evolve to the optimal case in which all the MSs send channel feedback in each frame.



Figure 4.4: System Goodput for available CQI slots with homogeneous users in (a) pedestrian mobility (b) vehicular mobility.

The results show that FEREP outperforms the reference scheme in both mobility scenarios. The performance gain is more significant in high mobility scenario and with most stringent total feedback budget (i.e., small F), which demonstrates that in such cases, using the same feedback strategy for all MSs is clearly not optimal, and that the proposed adaptive feedback

 $<sup>^{2}</sup>$ goodput = throughput - lossrate





Figure 4.5: (a) CQI (b) Feedback Window for a MS with ITU vehicular B channel model.

To get a more in-depth insight on the property of the proposed framework, we study the feedback window size at the output of the feedback window adaptation (FWA) module. Figure 4.5a plots the real channel condition from time 1 to 1000 of a vehicular MS and Figure 4.5b shows the correspondent feedback window size as derived by FWA. We can observe that with the average channel quality improvement over time, the average feedback window also increases. More specifically, when the channel quality varies significantly, the FWA module generates a relatively small feedback window, which allows the BS to have more channel feedback, when the channel quality variation decreases, the feedback window size increases as less feedback is required to estimate the channel condition. As a result, the FWA module dynamically tunes the feedback window size based on the channel condition, thus reaching an appropriate feedback window size.

The system parameter  $N_{ack}^{th}$  acts actively on the performance of FEREP scheme. Figure Figure 4.6 shows the results of FEREP with varying  $N_{ack}^{th}$  for VoIP and video applications for fixed mobility of 60 km/h. We observe that the average feedback window is inversely proportional to  $N_{ack}^{th}$  value. As  $N_{ack}^{th}$  increases, the system acts comparatively slowly to raise the value of w. Definitely, a smaller feedback window provides more knowledge about the channel state and helps to achieve higher goodput. Besides, larger feedback window saves uplink resources with the tradeoff of downlink throughput degradation. We found from our



Figure 4.6: Average feedback window for different  $N_{ack}^{th}$  values with homogeneous users in vehicular mobility.

simulation results that  $N_{ack}^{th} > 4$  maintains our target BER and PER.

#### 4.6.3 Heterogeneous Scenario

In this subsection, we study the case of a more realistic scenario including heterogeneous users. More specifically, 7 VoIP users and 7 video users are high priority users with service priority  $\alpha_k = 1.0$ , while the others are low priority users with  $\alpha_k = 0.5$ .  $g(\alpha_k)$  is set to a linear function  $g(\alpha_k) = \alpha_k$ . Figure 4.7 shows the system goodput as a function of F. We observe the same behavior as in the homogenous scenario, i.e., our proposed framework outperforms the reference scheme in both cases with more significant gain with high mobility channel and more stringent feedback budget.

We then study the average goodput of individual MS of different classes for VoIP and video applications in Figure 4.8 and Figure 4.9, respectively. We report the observation that the high priority users always enjoy better performance in terms of goodputs. Such performance gap between high priority and low priority users is more significant for video application in terms of goodputs. This can be explained as follows: as the packet size for VoIP is much smaller than that of video, the difference in the feedback window has much less impact on the final goodput for VoIP users than for video users. Indeed, we do observe significant difference in feedback window (thanks to the FWA module, high priority users have smaller



Figure 4.7: System goodput for available CQI slots with heterogeneous users in (a) pedestrian mobility (b) vehicular mobility.



Figure 4.8: System goodput for available CQI slots with heterogeneous VoIP users in (a) pedestrian mobility (b) vehicular mobility.

feedback window size, thus can send feedback more frequently), as shown in Figure 4.10. The above results demonstrate that the proposed framework is adapted to heterogeneous network scenarios and can be used easily to realize efficient service differentiation.



Figure 4.9: System Goodput for available CQI slots with heterogeneous Video users in (a) pedestrian mobility (b) vehicular mobility.



Figure 4.10: Average Feedback Window for available CQI slots in vehicular mobility with (a) VoIP users (b) Video users.

# 4.7 Conclusion

In this chapter, we have developed an integrated cross-layer framework of adaptive feedback with prediction for multiuser multicarrier OFDMA systems. The proposed framework, implemented at the BS side, is composed of three modules. The feedback window adaptation (FWA) module dynamically tunes the feedback window size for the mobile stations based on the received ack/nack that reflects their current channel condition. The priority-based feedback scheduling (PBFS) module then performs feedback scheduling by taking into account the feedback window size, the user profile and the total system feedback budget. The prediction based feedback (PBF) module performs CQI prediction by using recursive least square (RLS) algorithm for users which channel feedback that has not been scheduled in current frame. Through extensive simulations, the proposed framework shows significant performance gain especially in the case with stringent feedback budget constraint. Operators with 4th generation systems (i.e. Mobile WiMAX or LTE-Advanced) can easily take advantage of this framework to efficiently reduce feedback overhead while maintaining a high system performance in terms of radio resource utilization, service differentiation and adaptability.

However, having a strict feedback budget constraint in all frames forces some users to miss the chance of sending their CQI in time. To mitigate this issue, in next chapter, we propose to determine the number of CQI slots dynamically frame-by-frame. The ARQ protocol receives users acknowledgement only if the user is scheduled in the downlink. The reduction in users scheduling frequency also reduces the rate of ARQ hints and degrades the performance of above contributions. In this case, it is difficult to exploit the ARQ signal to adapt the feedback window for that user. In next chapter, we propose a dynamic feedback resource allocation (DCRA) scheme to address these issues.

# Chapter 5

# Dynamic CQI Resource Allocation for QoS Constrained OFDMA Systems

# 5.1 Introduction

In this chapter, we deal with the CQI allocation problem to reduce feedback overhead in a multi-user environment. Our main objective is to select a subset of users to send CQI among all the users to be scheduled in next frame while keeping the system throughput maximum and maintaining users' QoS constraints. To achieve this goal, we propose a dynamic CQI resource allocation (DCRA) algorithm to allocate resources for CQI feedback on frame-by-frame basis. The DCRA algorithm is a cross-layer scheme which uses the physical CQI information, ARQ protocol signals, Doppler based mobility tracking, and MAC layer service classifier. The use of a cross-layer strategy helps DCRA to combine information from two levels (physical and MAC) which jointly give a good insight on the channel quality of the user and its effect on the MAC frames in terms of PER (packet error rate), delay, etc. As each application reacts differently at MAC layer for the errors encountered at the physical layer, DCRA relies on hints from both MAC and physical levels to derive best feedback window periodicity to meet QoS application constraints.

# 5.2 System Model



Figure 5.1: The frame structure of the OFDMA/TDD system [29]

The BS allows F MSs to send feedback in each scheduling frame for  $0 \le F \le K$  and F is determined dynamically frame-by-frame to meet users QoS constraints. We assume uniform power allocation across users, thus the transmitted power is allocated identically to each subchannel. The system uses MCS, that is, the coding scheme and the modulation constellation is adapted to SNR realization of the selected MS.

From the known SNR  $\gamma_k(t)$  of user k, the user data rate at frame t is defined as (in terms of number of slots)

$$R_k(t) = \left\lceil \frac{L}{Q_i(\gamma_k(t)).N_{sym}} \right\rceil$$
(5.1)

where  $Q_i(\gamma_k(t))$  is the number of bits/symbol for SNR  $\gamma_k(t)$  with corresponding MCS level  $Q_i$ ;  $N_{sum}$  is the number of symbols per OFDMA slot; L is the data block size for a user.

# 5.3 Downlink Scheduling and Feedback Allocation Strategy

Due to the limited radio resources, OFDMA systems impose a budget constraint on the total number of feedback sent by the MSs to the BS. A critical question in this context is, under limited resource, which MSs should be admitted to send feedbacks. The goal in this work is to propose an adaptive feedback allocation strategy which purpose is two folds: i) allocating uplink feedback resources and ii) reducing total feedback overhead of the system without impacting the global QoS observed by users. Our feedback scheme operates in conjunction with the downlink scheduling module. Let us define the set of active users as  $\mathcal{K}$  with cardinality  $K = |\mathcal{K}|$ . In each frame a set of users  $\mathcal{S}, \mathcal{S} \subseteq \mathcal{K}$  (with cardinality S) are selected to be scheduled in downlink

and a set of users  $\mathcal{F}$  (with cardinality F),  $\mathcal{F} \subseteq \mathcal{K}$  are selected to send CQI feedback. As shown in Figure 5.1, at the beginning of downlink subframe (t-1), admissible users  $\mathcal{F}(t-1)$  will be informed to send CQI at uplink subframe (t-1). Once CQI feedbacks of  $\mathcal{F}(t-1)$  are received at uplink subframe (t-1), the PFS algorithm is performed to allocate resources to users in  $\mathcal{S}(t)$  for downlink transmission at frame t. The DCRA scheme determines  $\mathcal{F}(t)$  afterwards. The PFS and DCRA schemes are introduced in following sections.

#### 5.3.1 Downlink Proportional Fair Scheduling (PFS)

Schedulers in OFDMA-based systems can be designed to fully exploit multiuser and channel diversities in both time and frequency dimensions and to maximize total system capacity [72] [63] [81]. However, such schedulers do not take into account individual users' QoS. Other schemes have then been proposed to cope with users unequal channel conditions while keeping awareness of cell throughput maximization [70] [115] [106] [37] [71] [85] [51]. Among them, the proportional fair scheduling (PFS) scheme [70] has been very popular for its fairness among QoS constrained users. In this chapter, we adopt PFS scheme to schedule users in the downlink. We consider two main classes of users: realtime (RT) users which support latency-sensitive applications and non-realtime (NRT) users which are not time-constrained. To enable proportional fairness among all the users, we define  $U_k(t)$ , the priority value of MS k at frame t and  $U_k(t)$  is defined as

$$U_k(t) = \begin{cases} u_{rt} \cdot \{\frac{d_k(t)}{D_k(t)}\}, & \text{if } k \text{ is RT user}, \\ u_{nrt} \cdot \{\frac{r_k(t) - \bar{r_k}(t)}{r_k(t)}\}, & \text{if } k \text{ is NRT user} \end{cases}$$
(5.2)

where  $q_{rt}$  and  $q_{nrt}$  are the basic priorities for RT and NRT services, respectively and  $u_{rt} > u_{nrt}$ ;  $d_k$  is the delay that the head-of-line (HoL) packet of user k has experienced till frame t and  $D_k$ is the maximum delay tolerance;  $r_k$  is the minimum required transmission data rate of user k and  $\bar{r_k}(t)$  is the average transmission rate of user k till frame t. The  $\bar{r_k}(t)$  is given by

$$\bar{r}_k(t) = (1 - \frac{1}{T})\bar{r}_k(t-1) + \frac{1}{T}r_k(t), \quad k = 1, 2, ..., K$$
 (5.3)

where T is the observing window size. As OFDMA allows the BS to schedule multiple users at the same physical layer frame, the set of users S are selected to be scheduled in downlink of frame t with maximum priority score as

$$\mathcal{S}(t) = \arg \max_{\theta} \sum_{k \in \theta} U_k(t)$$
(5.4)

subject to

$$\sum_{k \in \theta} R_k(t) \le C_{dl}, \quad \forall R_k(t) > 0$$
(5.5)

where  $R_k(t)$  is the number of slots required by user k in frame t defined in Eq. 5.1,  $\theta$  denotes the possible subset of  $\mathcal{K} = \{1, 2, ..., K\}$ , and  $C_{dl}$  denotes the downlink capacity in number of slots. Therefore, there are  $2^K$  combination of  $\theta$ , which grows exponentially when K increases. For simplicity, we sort all users priority according to  $U_k(t)$ , then choose the  $S = |\mathcal{S}(t)|$  users with top priority to fit downlink capacity  $C_{dl}$  according to (5.5).

#### 5.3.2 CQI Feedback Scheduling Strategy



Figure 5.2: Decision making for feedback scheduling strategy for user k.

Let us define  $w_k$  as the average CQI feedback window of user k which means, user k sends CQI feedback once in every  $w_k$  frames. Now we define  $h_k$  as the average downlink scheduling window. It is determined by observing the downlink scheduling frequency of user k (expressed in number of frames) and is used as an input to determine our feedback scheduling strategy as shown in Figure 5.2.

#### **Deterministic Feedback Scheduling**

When  $h_k$  presents large values, user k is not scheduled in downlink very frequently and therefore does not require frequent feedback in uplink. The BS specifically asks to send feedback only those users who are going to be scheduled in the next frame. We call this method as deterministic feedback scheduling (DFS). With this method, no further feedback optimization is possible. A new and fresh feedback is mandatory since the last feedback value is out of date and cannot be exploited (large  $h_k$  value) and this falls in the case of  $(h_k \ge w_k)$ . In other notation, we say  $S \subseteq \mathcal{K}$  and  $S = \mathcal{F}$ .

#### **Opportunistic Feedback Scheduling**

In this scheme, the BS broadcasts the CQI threshold  $\gamma_{th}$  and the users  $\mathcal{F} = \{k | \gamma_k > \gamma_{th}, k = 1, 2, ..., K\}$  send the feedback to BS. Then the BS selects  $\mathcal{S} \subseteq \mathcal{F}$  users to schedule in the downlink using PFS. We call this CQI feedback method as opportunistic feedback scheduling (OFS). There is a CQI feedback outage if  $\mathcal{S} \subsetneq \mathcal{F}$  and the BS randomly selects some users to schedule in downlink using the most robust MCS level. To avoid the feedback outage problem, the BS needs to rebroadcast the CQI threshold  $\gamma_{th}$  with a smaller value. This also increases the opportunity for more users to send CQI feedback hence increases feedback overhead.

As the number of high data rate users increases (more specially realtime users), BS needs to schedule them more frequently to fulfill their delay constraints  $D_k$  (expressed in terms of number of frames), hence leading to users with smaller scheduling window  $h_k$ . To capitalize multiuser diversity and exploit the MCS, the BS also requires more frequent feedbacks from those users leading to smaller  $w_k$  values. In this case of high feedback load, DCRA scheme is used to reduce the feedback resources.

### 5.4 Dynamic CQI Resource Allocation (DCRA)

DCRA is the core part of our feedback allocation strategy. In DCRA, to achieve feedback overhead reduction with minimum system throughput degradation, the base station adapts the feedback window of each user dynamically to fulfill its applications QoS requirements. MSs may have different types of applications (realtime, non-realtime) and mobility profiles (e.g. vehicular and pedestrian). To maintain QoS demands (for example in terms of delay or loss rate), we propose to use MAC layer hints to adapt feedback window as these hints provide a good global insight about the channel condition effect on MAC frames. As previously, we assume an ARQ protocol with go-back-n mechanism. The performance of ARQ protocol is highly dependent on users applications and mobility profiles. As an example, packet size of video traffic is usually larger than VoIP traffic. The packet error rate (PER) varies over packet size for the same target BER. On the other hand, PER varies over BER for same target packet size. Sending an ack/nack of ARQ protocol depends on the PER of the packet. This PER diversity among different application and mobility users can be used to adapt their feedback rate to achieve maximum system throughput. We note that similar approach of using ARQ signaling has been proposed in [62][111] to solve other networking issues like power and rate adaptation. The authors used ack/nack signals to cope up with imperfect CQI feedbacks in non-constraint feedback budget environment (i.e. F = K).

A necessary condition for the high-performance of DCRA is the availability of frequent ack/nack hints. Such ARQ signaling rate is strongly related to the scheduling window  $h_k$ . Hence, based on the scheduling window of a user, we propose two operation modes for DCRA: i) DCRA with realtime feedback window adaptation (RTw-DCRA), and ii) DCRA with empirical feedback window (EMw-DCRA). In case of frequent user scheduling (for example  $h_k = 1$ , i.e. user is scheduled in every frame), frequent ARQ signaling can compensate the lack of frequent CQI feedback. Such knowledge allows to further reduce the feedback overhead using larger  $w_k$  as long as the MAC layer is not affected by errors. RTw-DCRA adapts feedback window in realtime which allows the BS to determine average feedback window size on a frame-byframe basis for all users. EMw-DCRA uses the empirically determined feedback windows to schedule feedbacks while the ARQ signaling is not frequent because the user is not scheduled very frequently (e.g. in each frame) in downlink.

#### 5.4.1 RTw-DCRA: DCRA with Realtime Feedback Window Adaptation

RTw-DCRA part of DCRA is executed in realtime for those users which show a small scheduling window size  $h_k$ . The objective of RTw-DCRA algorithm is to dynamically calculate the appropriate feedback window size based on MSs' channel condition. We compute the feedback window for each user based on the feedback window adaptation (FWA) algorithm proposed in chapter 4. The algorithm is shown again in Algorithm 5.

We take benefit of this feedback window adaptation procedure in the DCRA algorithm shown in Algorithm 6. DCRA determines the set of users  $\mathcal{F}$  to send feedback in each frame among all the users selected to be scheduled in downlink (set  $\mathcal{S}$ ) by PFS. The feedback window

```
Algorithm 5 Feedback Window Adaptation procedure to calculate w_k
```

```
1: procedure Adapt_w_k
```

- 2: input:  $N_{ack}^{th}, w_k$
- 3: Set  $n_{ack} \leftarrow 0$ ,
- 4: if ack received then
- 5:  $n_{ack} \leftarrow n_{ack} + 1$
- 6: **if**  $n_{ack} \ge N_{ack}^{th}$  then
- 7:  $w_k \leftarrow w_k + 1$
- 8:  $n_{ack} \leftarrow 0$
- 9: **end if**
- 10: else if nack received or timeout then

11: 
$$w_k \leftarrow \max\{w_k - 1, 1\}$$

```
12: n_{ack} \leftarrow 0
```

- 13: end if
- 14: return  $w_k$

of each user k is maintained separately by BS and initialized with 1 (line 3). The user k is scheduled to send feedback if the BS did not receive feedback since  $w_k$  frames (lines 14 – 19). Each time a user k is scheduled, we calculate its average scheduling window  $h_k$  (line 17). If  $h_k$ is sufficiently small so that we have enough ARQ hints, procedure Adapt\_w\_k is called (lines 7-8) to adapt  $w_k$  accordingly. We define a threshold  $h_{th}$  to decide how much small the  $h_k$ should be in order to apply the procedure Adapt\_w\_k.  $h_{th}$  is a system parameter to be tuned by system operator.

For some user k, if  $h_k > h_{th}$  then the BS may not have enough ARQ hints to decide the feedback window perfectly for that user. In this case, the BS depends on a empirically determined feedback window matrix W. Each entry of W is a function of service class e and moving velocity v. The system parameter  $h_{th}$  determines when the BS should use RTw-DCRA or EMw-DCRA. If the  $h_k$  is larger than  $h_{th}$  but smaller than  $w_k$ , the BS determines the service class e and mobility profile v for the user and uses the feedback window with W(e, v) value (lines 9 – 12). The procedure to determine matrix W is explained in next subsection.

Algorithm 6 Dynamic CQI Resource Allocation procedure 1: procedure DCRA 2: input: W, T3: Set  $t \leftarrow 0, t_k^{last} \leftarrow 0, h_k \leftarrow 0, w_k \leftarrow 1$ 4: **loop**  $\mathcal{F} \leftarrow \{\}$ 5: for each user  $k \in S$  to be scheduled **do** 6: if  $h_k < h_{th}$  then 7:  $w_k \leftarrow \text{procedure Adapt_w_k}$ 8: else if  $h_k < w_k$  then 9:  $e \leftarrow \text{determine service class of user } k$ 10:  $v \leftarrow$  determine mobility profile of user k 11:  $w_k \leftarrow W(e, v)$ 12:end if 13:if  $t - t_k^{last} \ge w_k$  then 14:schedule user k to send feedback 15: $\mathcal{F} \leftarrow \mathcal{F} \cup \{k\}$ 16: $h_k \leftarrow (1 - \frac{1}{T})h_k + \frac{1}{T}(t - t_k^{last})$ 17: $t_k^{last} \gets t$ 18:end if 19:end for 20: allocate  $|\mathcal{F}|$  slots in uplink for CQI feedback 21: $t \leftarrow t+1$ 22:23: end loop

| Algorithm 7 Procedure to determine matrix $W$                               |  |  |  |  |
|---|--|--|--|--|
| 1: procedure DetermineW   |  |  |  |  |
| 2: for each $e \in \mathcal{E}$ do  |  |  |  |  |
| 3: for each $v \in \mathcal{V}$ do  |  |  |  |  |
| 4: Set $N_{ack}^{th}, w_k \leftarrow w_{init}, w_{avg} = 0, i \leftarrow 0$ |  |  |  |  |
| 5: <b>loop</b>  |  |  |  |  |
| 6: $w_{saved}(i) \leftarrow \text{procedure Adapt_w_k}$                     |  |  |  |  |
| 7: $i \leftarrow i+1$   |  |  |  |  |
| 8: $w_{avg} \leftarrow \text{average}(w_{saved})$                           |  |  |  |  |
| 9: end loop   |  |  |  |  |
| 10: $W(e,v) \leftarrow w_{avg}$   |  |  |  |  |
| 11: end for   |  |  |  |  |
| 12: end for   |  |  |  |  |
| 13: end procedure   |  |  |  |  |
|   |  |  |  |  |

#### 5.4.2 EMw-DCRA: DCRA with Empirical Feedback Window

During the network planning phase, the network operator creates a feedback window matrix W for each cell in the network by running extensive experiments according to our Algorithm 7. The operator first determines its supported service classes and mobility profiles. Each service class has specific target BER, PER, packet size, delay constraints which are determined according to users QoS requirements. As small velocity variation does not change the variance of Doppler spread significantly, the change in feedback rate also would not be significant. Hence, we define a finite number of mobility profiles, each one supporting a range of moving velocities. Let us define the set of all the possible service classes as  $\mathcal{E}$  and the set of all possible mobility profiles as  $\mathcal{V}$ . The network operator then launches several experiments with different values of e and v to derive the corresponding average feedback window w. It is worth mentioning that in these experiments users evolve in the same cell propagation environment and that inter-cell interference is neglected.

To determine the average feedback window w for each e and v (lines 2-3), initially w is set to  $w_{init}$  (line 4). Then, an iteration loop executes the feedback window adaptation procedure Adapt\_w\_k (line 6). In each iteration, Adapt\_w\_k returns the adapted window w which is stored in array  $w_{saved}$  to calculate the average  $w_{avg}$  (lines 7 - 8). Finally,  $w_{avg}$  is saved in W(e, v)(line 10). The experiment duration to determine each W(e, v) entry should be long enough to ensure that  $w_{avg}$  is sufficiently steady. In our simulations, we found that  $w_{avg}$  converges within  $10^{-4}$  after 10000 iterations.

| Parameters                    | Value  |  |
|-------------------------------|--|--|
| Channel bandwidth             | 10 MHz                                       |  |
| Frame duration                | $5 \mathrm{ms}$                              |  |
| Downlink slot                 | PUSC (1 channel $\times$ 2 symbols)          |  |
| Uplink slot                   | PUSC (1 channel $\times$ 3 symbols)          |  |
| # of downlink slots $C_{dl}$  | 330  |  |
| # of uplink slots $C_{ul}$    | 210  |  |
| Cell radius                   | 1 km   |  |
| Path loss PL (d)              | $12 log_{10}(4\pi d/\lambda) - 27$           |  |
| Video packet size per frame   | 121 bytes                                    |  |
| VoIP packet size per frame    | 12 bytes                                     |  |
| Mobility model                | ITU Vehicular with 3, 15, 30, 45, 60 km/h $$ |  |
| ARQ maximum number of retries | 4  |  |
| $N^{th}_{ack}$                | 4  |  |
| Target BER                    | $10^{-3}$                                    |  |
| Target packet loss rate       | 1%   |  |
|                               | 1000   |  |
| $w_{init}$                    | 1  |  |

Table 5.1: Simulation Parameters

### 5.5 Simulation Results and Analysis

In this section, we evaluate the performance of the proposed DCRA scheme via extensive simulations and gain further insight on how different system parameters influence the performance. The simulations are performed using Matlab tool [83]. We study the behavior of DCRA in both empirical feedback window adaptation (EMw-DCRA) and realtime feedback window adaptation (RTw-DCRA) scenarios. For each scenario, we conduct a set of simulations

| Modulation | Level | Coding | Minimum Receiver SNR (db) |
|------------|-------|--------|---------------------------|
| BPSK       | 1     | 1/2    | 3.0                       |
| QPSK       | 2     | 1/2    | 5.5                       |
|            | 3     | 3/4    | 8.0                       |
| 16QAM      | 4     | 1/2    | 10.0                      |
|            | 5     | 3/4    | 13.5                      |
| 640AM      | 6     | 2/3    | 17.0                      |
| 04QAINI    | 7     | 3/4    | 18.5                      |

Table 5.2: Minimum receiver SNR required to use in MCS and their bitrates per symbol per carrier

using 5 channel patterns indicative of the typical Mobile WiMAX mobility profiles, from the ITU pedestrian A model representing low mobility scenario of 3km/h to the ITU vehicular B model corresponding to high mobility scenario of 60km/h. Then we compare the performance of DCRA with two reference mechanisms (i) DFS the deterministic feedback scheduling proposed in many works [89][86] [64], (ii) OFS the opportunistic feedback scheme introduced in [46][107]. For downlink resource allocation, all three schemes use similar proportional fair scheduling (PFS) as in [70].

#### 5.5.1 Simulation Environment

We study a single-cell, single-sector system where the BS communicates with upto 32 MSs randomly distributed in the cell with radius 1km. The network parameters, as shown in Table 5.1, are set following the IEEE 802.16e standard [89]. As mentioned previously, 5 channel mobility models are simulated: ITU pedestrian A (3km/h), 15km/h, 30km/h, 45km/h and ITU vehicular B (60km/h). The system supports two service classes with video and VoIP applications. Video and VoIP users have a data rate of 176 kbps with H.264 format and 5.3 kbps with G723.1 [101] respectively. Taking consideration of MAC header and fragmentation or packing header overhead, the data rates are 189 kbps and 18.75 kbps for video and VoIP users, respectively. During a simulation scenario, 50% of the total users have service class of video application and rest 50% are VoIP users. The minimum received SNR required to decide each MCS level is shown in Table 5.2.

In the simulation scenario, MSs are created randomly in the cell and use simplified random waypoint model to move within the cell with different constant velocities. Application session duration is fixed to 200 seconds which is equivalent to 40000 frames. Each simulation is run 300 times with different seeds then we calculate the means and 95% confidence intervals to plot the curves.

#### 5.5.2 Empirical Study: Determining Matrix W



Figure 5.3: Average feedback window for different vehicular mobility speeds.

To evaluate the performance of DCRA, first we derive the matrix W. We simulate using an application delay constraint  $D_k = 1$  frame  $\forall k$  so that the average scheduling window becomes  $h_k = 1$ . By doing so, we can receive the ARQ hints in frame-by-frame basis and fully exploit this cross-layer knowledge. In realtime operations, if we have less ARQ hints in the case of large  $h_k$  values, we use W to take intuitive decisions to choose  $w_k$  for a user with a known service class and mobility profile. In Figure 5.3, we present the average feedback window experienced by a user for different service classes and for each mobility profile. The figure shows that the average feedback window decreases with the increase of user mobility. High mobility users require to send feedbacks more frequently to track their large channel variation. It also shows different feedback windows for video and VoIP applications due to packet size.

As we have stated in section 5.4.1, the system parameter  $N_{ack}^{th}$  acts actively on the performance of DCRA scheme. Figure 5.4 shows the results of DCRA scheme with varying  $N_{ack}^{th}$  for video and VoIP applications for fixed mobility of 60 km/h. We observe that the average feed-



Figure 5.4: Average feedback window for different  $N_{ack}^{th}$  values.

back window is inversely proportional to  $N_{ack}^{th}$  value. As  $N_{ack}^{th}$  increases, the system acts more lately to raise the value of w. Definitely, a smaller feedback window provides more knowledge about the channel state and helps to achieve higher throughout. Besides, larger feedback window saves uplink resources with the tradeoff of downlink throughput degradation. We found from our simulation results that  $N_{ack}^{th} \geq 4$  maintains our target BER and PER.

#### 5.5.3 Performance of DCRA

In this set of experiments, we set CQI threshold  $\gamma_{th} = 10$ db for OFS scheme and maintained it fixed for all delay constrained applications even though it is possible to change this threshold value in realtime. Note that in terms of overhead, the opportunistic approach requires additional signaling to announce the new threshold values. In our simulations, we neglected the protocol signaling overhead and only consumed resources for CQI slots have been considered. Figure 5.5 shows the average throughput per user as a function of D (the delay constraint of realtime applications). It shows that our scheme always outperforms OFS scheme but presents lower performance (maximum 7% degraded throughput for  $D_k = 3$ ) compared to the DFS, in case of low delay constrained applications. This is mainly due to the reduced feedback of DCRA which involves less knowledge about channel condition. The performance of DCRA converges to DFS for high delay constraints as  $h_k > w_k$ , while OFS shows lower performance



Figure 5.5: Average throughput per user for varying application delay constraint



Figure 5.6: Average packet loss rate per user for varying application delay constraint

because of its large outage probability. According to the OFS scheme specifications, the BS transmits data in downlink with a robust MCS level if there is a feedback outage for a scheduled user.

Figure 5.6 shows the packet loss rate as a function of D. For low delay constrained applications, DCRA under performs compared to DFS and OFS but the loss rate never exceeds 3.5%



Figure 5.7: Average CQI feedback load for varying application delay constraint



Figure 5.8: Average throughput per user for varying number of active users

which is tolerable for most realtime applications (e.g. VoIP and video). Figure 5.7 presents the average feedback overhead per frame for different delay limits and shows that DCRA always outperforms DFS and OFS. In this scenario, DCRA reduces up o 37% feedback load compared to DFS and up to 400% compared to OFS while maintaining the QoS constraints (target delay and lossrate) of the applications. One may argue that the feedback overhead for OFS can be



Figure 5.9: Average packet loss rate per user for varying number of active users



Figure 5.10: Average CQI feedback load for varying number of active users

further reduced by increasing the CQI threshold but this will be at the expense of increased feedback outage probability degrading the overall system throughput. Note that the feedback overhead with DCRA is lower than with DFS for  $D \leq 6$ . For D > 6, DCRA scheme has  $h_k > w_k$  and converges to DFS performance.

Finally, we study the system performance while varying K, the number of active users in

the network. We increase K while keeping the total system downlink capacity  $C_{dl}$  fixed to 330 OFDMA slots and D = 4. Figure 5.8 shows the average throughput per user. All schemes report similar performance for limited number of users ( $k \leq 20$ ). The average throughput of DCRA as well as reference schemes decreases with increasing K. However, our scheme still outperforms OFS scheme. Figure 5.9 shows the average packet loss rate in the network which reports similar performance compared to the reference schemes. Figure 5.10 presents the feedback overhead for different values of total number of users. It proves that DCRA scheme can maintain downlink system performance similar to other schemes while reducing the feedback overhead significantly and maintaining QoS requirements of each user.

### 5.6 Conclusion

In this chapter, we proposed a cross-layer dynamic CQI resource allocation (DCRA) algorithm for multiuser multicarrier OFDMA systems. DCRA uses the ARQ information, Doppler based mobility tracking, and MAC layer service classifier. During the network planning phase, an operator can derive the average feedback window values based on the user velocity and service class for a target packet error rate. Our experimental analysis shows that the feedback window can be averaged according to users service class and their mobility profile for a given cell environment. The empirical information collected during network planning phase can be extremely valuable for network operator to achieve the dynamic CQI resource allocation during realtime operation. DCRA performs a realtime dynamic window adaptation in case of sufficient cross-layer hints are available from ARQ signaling. DCRA increases uplink resource by reducing feedback overhead without degrading downlink throughout significantly compared to deterministic feedback scheduling (DFS) and opportunistic feedback scheduling (OFS). From the users perspective, DCRA improves QoS constraints like packet loss rate and saves MS power due to feedback reduction.

# Chapter 6

# Conclusions

In this chapter, we first conclude our work on efficient CQI feedback resource allocation, and discuss the advantages and disadvantages of CQI feedback resource allocation. We then present some open problems in this research area.

# 6.1 Conclusions

The wireless communications industry is gaining momentum in both fixed and mobile applications. The continued increase in demand for all types of wireless services (voice, data, and multimedia) is fueling the need for higher capacity and data rates not only in fixed but also in mobile applications. WLANs and 3G cellular networks are experiencing several difficulties for reaching a complete mobile broadband access, bounded by factors such as bandwidth, coverage area, or infrastructure costs. In this context, multiuser multicarrier OFDMA systems like Mobile WiMAX, LTE-Advance appear to fulfil these requirements, providing vehicular mobility and high service areas and data rates. Defined as broadband wireless access, they are increasingly providing high data rate applications like video streaming, group video chatting, online gaming, etc. Since the channel condition of a mobile station (MS) changes over time due to the fading, interference and path loss, the wireless systems use adaptive modulation and coding scheme (MCS) techniques to change their sending rates based on the CQI in terms of estimated signal-to-noise ratio (SNR). If the base station receives CQI feedbacks from all users, large feedback overhead exhaust the limited uplink capacity. This thesis is devoted to the study of CQI feedback resource allocation for these multiuser multicarrier OFDMA systems. In chapter 2, we provided a literature review related to the feedback issues in OFDMA systems. At first, we introduced the OFDMA technology, then we presented the CQI basics and main feedback mechanisms. We described the related works dealing with feedback reduction and feedback resource allocation. Afterwards, we review the state of the art in channel predictions.

In chapter 3, we reviewed a model for multiuser multicarrier OFDMA systems together with a model for random time varying wireless channel. We also reviewed the CQI generation and CQI quantization model. We reviewed the doppler mobility model and ARQ protocol to use in our cross-layer approaches. These models provided a basis for our development in subsequent chapters. We presented the most used periodic feedback mechanism for distributed subcarriers as interval based feedback (IBF). Then we proposed prediction based feedback (PBF) which allows the BS to predict the CQI feedbacks based on RLS algorithm [14]. The PBF reduces the error variance compared to interval based feedback (IBF) and reduces over and under estimation variations. We showed that it is useful to use channel prediction as a tool to reduce the feedback overhead and improve the uplink throughput.

Then we introduced the analytical model for system performance with and without prediction [13]. We defined the error distribution from the error data to derive the probability of error in each adaptive modulation level and some performance metrics with both predicted CQI and actual CQI. Our analytic model and numerical results showed the throughput degradation in downlink channel created by the lack of CQI feedback. The performance degradation becomes negligible for channels with high SNR. In order to mitigate the possible under and over estimation effects of CQI prediction, we proposed an opportunistic periodic feedback mechanism [18]. In this mechanism, we exploited the hints from MAC-layer ARQ protocol to enhance the performance of periodic feedback mechanisms. The opportunistic mechanism improves the system performance for high mobility cases compared to low mobility cases.

In chapter 4, we have developed an integrated cross-layer framework for feedback resource allocation and prediction (FEREP) for multiuser multicarrier OFDMA systems [19]. The proposed framework, implemented at the BS side, is composed of three modules. The feedback window adaptation (FWA) module dynamically tunes the feedback window size for the mobile stations based on the received ack/nack that reflects their current channel condition [15]. In current systems, the operators may reserve a fixed number of slots for CQI feedback which is not changed frame-by-frame. If the number of users is greater than number of CQI slots, users compete for the CQI resource. To cope with this issue, we proposed priority-based feedback scheduling (PBFS) module which performs feedback scheduling by taking into account the feedback window size, the user profile and the total system feedback budget.

The prediction based feedback (PBF) module performs CQI prediction by using recursive least square (RLS) algorithm for user which channel feedback has not been scheduled in current frame. Through extensive simulations, the proposed framework shows significant performance gain especially in the case with stringent feedback budget constraint. Operators exploiting 4G systems (i.e. Mobile WiMAX or LTE-Advanced) can easily take advantage of this framework to efficiently reduce feedback overhead while maintaining a high system performance in terms of radio resource utilization, service differentiation and adaptability.

In chapter 5, we proposed a cross-layer dynamic CQI resource allocation (DCRA) algorithm for multiuser multicarrier OFDMA systems [16] [17]. While FEREP operates on a strict feedback budget constrained scenario, DCRA operates on a relaxed CQI feedback budget in a frame-by-frame basis to determine the necessary CQI budget exploiting the cell dynamics. DCRA uses two modes for feedback window estimation. The first one is an off-line mode based on empirical studies to derive optimal average feedback window based on user application and mobility profile. ARQ protocol receives users acknowledgement only if the user is scheduled in the downlink. The reduction in users scheduling frequency also reduces the rate of ARQ hints and degrades the performance of above contributions. In this case, it is difficult to exploit the ARQ signal to adapt the feedback window for that user. During the network planning phase, an operator can derive the average feedback window values based on the user velocity and service class for a target packet error rate. Our experimental analysis showed that the feedback window can be averaged according to users service class and their mobility profile for a given cell environment.

The empirical information collected during network planning phase can be extremely valuable for network operator to achieve the dynamic CQI resource allocation during realtime operation. DCRA performs a realtime dynamic window adaptation in case of sufficient crosslayer hints are available from ARQ signaling. DCRA increases uplink resource by reducing feedback overhead without degrading downlink throughout significantly compared to deterministic feedback scheduling (DFS) and opportunistic feedback scheduling (OFS). From the users perspective, DCRA improves QoS constraints like packet loss rate and saves MS power due to feedback reduction.

### 6.2 Issues for Further Work

We finally provide some suggestions for future research concerning possible extensions of the work presented in this thesis.

- The idea of CQI feedback prediction can be extended to wireless MIMO-OFDMA systems which have been proposed for modern 4G communication systems. It has been demonstrated in [53] that CQI is required to exploit the full potential of MIMO systems. This CQI can be acquired by prediction. An important application of such a scheme could be space-time decoding. If the spatial correlations or structure of the MIMO channel can be exploited, the performance of channel prediction can be expected to improve. However, the results will critically depend on the channel model used.
- The idea of CQI feedback prediction for distributed subcarriers can be extended to the contiguous subcarriers like band-AMC. But it will also increase the space and time complexity of the system.
- Our theoretical analysis of prediction error could be extended to communication systems using orthogonal frequency division multiple access (OFDMA) for the uplink. Specifically, it would be interesting to consider the capacity regions for OFDMA communications over time and frequency selective fading channels.
- As regards to the prediction tool, we used the feedback prediction model using RLS which has certain prediction errors. These errors create over and under estimations of the channel. One can reduce these estimation errors by improving the prediction algorithm.
- With the integration of MIMO features, the feedback overhead increases drastically with the number of transmitting and receiving antenna. An essential investigation can be the adaptation of FEREP or DCRA schemes in open loop and closed loop MIMO scenarios. In the open loop MIMO, it is required for each MS to send only one CQI value (the best one) instead of whole CQI matrix. Our ongoing investigations show that the application of FEREP or DCRA in open loop MIMO is more intuitive than close loop MIMO.
- We have used the cross-layer information from MAC-layer ARQ protocol to enhance the performance of our schemes. We also used a decision making parameter to determine to use the offline or online mode of DCRA scheme. We could extend the idea of cross-layer

mechanism to use application layer information, more specifically, for high datarate video applications with on-off model of data bursts. One idea could be that the BS schedules the MSs to send CQI feedback during their on period of data bursts, and MSs with no data burst do not send any feedback.

• Similar to the majority of researches in OFDMA systems, we have used the simulation as a tool to evaluate the performance of our approaches and to extrapolate conclusions. It would be useful to verify the simulation results by implementing our FEREP and DCRA approaches and measure the performance in a practical 4G network.

# Bibliography

- 3rd Generation Partnership Project (3GPP). Physical layer aspect for evolved universal terrestrial radio access (utra). Technical report, European Telecommunications Standards Institute (ETSI), 2006.
- [2] IEEE 802.11a/D7.0 1999. Part 11: Wireless lan medium access control (mac) and physical layer (phy) specifications: High speed physical layer in the ghz band, 1999.
- [3] Rajiv Agarwal, V. Abhishek, R. Vannithamby, and J. Cioffi. Opportunistic feedback in ofdma downlink. In *IEEE Global Telecommunication Conference*. IEEE, 2007.
- [4] Rajiv Agarwal, V. Majjigi, Z. Han, R. Vannithamby, and J. Cioffi. Low complexity resource allocation with opportunistic feedback over downlink ofdma networks. *IEEE Journal on Selected Areas in Communications*, 26(8):1462–1472, 2008.
- [5] Rajiv Agarwal, Vinay Majjigi, Rath Vannithamby, and John Cioffi. Efficient scheduling for heterogeneous services in ofdma downlink. In *IEEE Global Telecommunications Conference (GLOBECOM'07)*. IEEE, November 2007.
- [6] Rajiv Agarwal, Rath Vannithamby, and John M. Cioffi. Optimal allocation of feedback bits for downlink ofdma systems. In *International Symposium on Information Theory*. IEEE, July 2008.
- [7] Rohit Aggarwal, Mohamad Assaad, C. Emre Koksal, and Philip Schniter. Ofdma downlink resource allocation via arq feedback. In 43rd Asilomar conference on Signals, systems and computers. IEEE, November 2009.
- [8] Yahya S. Al-Harthi, Ahmed H. Tewfik, and Mohamed-Slim. Alouini. Multiuser diversity with quantized feedback. *IEEE Transactions on Wireless Communications*, 6(1):330– 337, January 2007.

- [9] M.-S. Alouini and A.J. Goldsmith. Adaptive modulation and nakagami fading channels. *Kluwer Journal of Wireless Communications*, 13(1-2):119–145, May 2000.
- [10] J. B. Andersen, J. Jensen, S.H. Jensen, and F. Frederiksen. Prediction of future fading based on past measurements. In *IEEE Vehicular Technology Conference (VTC'99-Fall.* IEEE, 1999.
- [11] Daniel Aronsson and Mikael Sternad. Kalman predictor design for frequency-adaptive scheduling of fdd ofdma uplinks. In *IEEE International Symposium on Personal, Indoor* and Mobile Radio Communications (PIMRC). IEEE, February 2007.
- M. A. Awal and L. Boukhatem. Wimax and end-to-end qos support. In Radio Resources Management in WiMAX: From theoretical capacity to system simulations, pages 105–138.
   ISTE & Wiley, 2009.
- [13] Mohammad Abdul Awal and Lila Boukhatem. Analysis of feedback prediction error on the downlink performance of ofdma systems. In 20th Personal, Indoor and Mobile Radio Communications Symposium (PIMRC'09). IEEE, September 2009.
- [14] Mohammad Abdul Awal and Lila Boukhatem. Effect of feedback prediction on ofdma system throughput. In 5th International Wireless Communications and Mobile Computing Conference (IWCMC'09). ACM, June 2009.
- [15] Mohammad Abdul Awal and Lila Boukhatem. Cross-layer cqi feedback window adaptation for ofdma systems under feedback budget constraint. In *IEEE Symposium on Computers and Communications (ISCC'11)*. IEEE, June 2011.
- [16] Mohammad Abdul Awal and Lila Boukhatem. Dynamic cqi resource allocation for ofdma systems. In *IEEE Wireless Communication and Networking Conference (WCNC'11)*. IEEE, March 2011.
- [17] Mohammad Abdul Awal and Lila Boukhatem. Efficient cqi resource allocation for qos constrained ofdma systems. *journal*, 2011. under preparation.
- [18] Mohammad Abdul Awal and Lila Boukhatem. Opportunistic periodic feedback mechanisms for ofdma systems under feedback budget constraint. In *IEEE Vehicular Technol*ogy Conference (VTC'11-Spring). IEEE, May 2011.

- [19] Mohammad Abdul Awal, Lila Boukhatem, and Lin Chen. An integrated cross-layer framework of adaptive feedback resource allocation and prediction for ofdma systems. *Journal of Computer Networks*, 2011. Submitted.
- [20] Krishna Balachandran, Doru Cailin, Niranjan Sudhir Jishi, Joseph H. Kang, and Achilles Kogiantis. Method for resource allocation in a wireless communication system, March 2008.
- [21] Hans Jørgen Bang, Torbjörn Ekman, and David Gesbert. Channel predictive proportional fair scheduling. *IEEE Transactions on Wireless Communications*, 7(2):482–487, February 2008.
- [22] J. A. C. Bingham. Multicarrier modulation for data transmission An idea whose time has come. *IEEE Communication Magazine*, 28:5–14, May 1990.
- [23] D.T. Chen. On the analysis of using 802.16e wimax for point-to-point wireless backhaul. In *Radio and Wireless Symposium*. IEEE, January 2007.
- [24] Jieying Chen, Randall A. Berry, and Michael L. Honig. Performance of limited feedback schemes for downlink ofdma with finite coherence time. In *International Symposium on Information Theory.* IEEE, June 2007.
- [25] Jieying Chen, Randall A. Berry, and Michael L. Honig. Limited feedback schemes for downlink ofdma based on sub-channel groups. *IEEE Journal on Selected Areas in Communications*, 26(8):1451–1461, 2008.
- [26] Y.-J. Choi and S. Bahk. Partial channel feedback schemes maximizing overall efficiency in wireless networks. *IEEE Transactions on Wireless Communications*, 7(4):1306–1314, April 2008.
- [27] Y. J. Choi, J. Kim, and S. Bahk. Qos-aware selective feedback and optimal channel allocation in multiple shared channel environments. *IEEE Transactions on Wireless Communications*, 5(11):1–9, 2006.
- [28] S.T. Chung and A.J. Goldsmith. Degrees of freedom in adaptive modulation: A unified view. *IEEE Transactions on Communications*, 49(9):1561–1571, September 2001.
- [29] Yao-Hsing Chung and Chung-Ju Chang. Opportunistic scheduling with economized csi feedback for ofdma/tdd downlink systems. In 20th Personal, Indoor and Mobile Radio Communications Symposium (PIMRC'09). IEEE, September 2009.

- [30] L.J. Cimini. Analysis and simulation of a digital mobile channel using orthogonal frequency division multiplexing. *IEEE Transactions of Communication*, 33(7):665–675, June 1985.
- [31] Reuven Cohen and Guy Grebla. Efficient allocation of CQI channels in broadband wireless networks. In *Infocom'2011 (mini-conference)*. ACM, April 2011.
- [32] M. Dianati and R. Tafazolli. Opportunistic scheduling over wireless fading channels without explicit feedback. In Vehicular Technology Conference, VTC2008-Spring. IEEE, May 2008.
- [33] A. Duel-Hallen, H. Hallen, and T.-S. Yang. Long range prediction and reduced feedback for mobile radio adaptive ofdm systems. *IEEE Transaction on Wireless Communication*, 5(10):2723–2733, October 2006.
- [34] A. Duel-Hallen, S. Hu, and H. Hallen. Long range prediction of fading signals: Enabling adaptive transmission for mobile radio channels. *IEEE Signal Processing Magazins*, 17(3):6275, May 2000.
- [35] T. Ekman and G. Kubin. Nonlinear prediction of mobile radio channels: Measurements and mars modeling designs. In *IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, March 1999.
- [36] Torbjorn Ekman. Prediction of Mobile Radio Channels. PhD thesis, Uppsala University, 2002.
- [37] M. Ergen, S. Coleri, and P. Varaiya. Qos aware adaptive resource allocation techniques for fair scheduling in ofdma based broadband wireless access systems. *IEEE Transactions* on Broadcasting, 49(4):362–370, 2003.
- [38] Thomas Eriksson and Tony Ottosson. Compression of feedback for adaptive transmission and scheduling. *IEEE*, 95(12):2314–2321, 2007.
- [39] European Telecommunications Standard Institute (ETSI). Digital audio broadcasting (dab) to mobile, portable and fixed receivers. Technical Report TS 300 401, European Telecommunications Standards Institute (ETSI), 1995.
- [40] European Telecommunications Standard Institute (ETSI). Digital video broadcasting (dvb); framing structure, channel coding and modulation for digital terrestrial television.

Technical Report TS 300 744, European Telecommunications Standards Institute (ETSI), 1999.

- [41] T. Eyceoz, A. Duel-Hallen, and H. Hallen. Deterministic channel modeling and long range prediction of fast fading mobile radio channels. *IEEE Communications Letters*, 2(9):254256, September 1998.
- [42] T. Eyceoz, S. Hu, and A. Duel-Hallen. Performance analysis of long range prediction for fast fading channels. In 33rd Conference on Information Sciences and Systems, March 1999.
- [43] S. Falahati, A. Svensson, T. Ekman, and M. Stenard. Adaptive modulation systems for predicted wireless channels. *IEEE Transactions on Communications*, 52(2):307316, February 2004.
- [44] F. Floren, O. Edfors, and B.-A. Molin. The effect of feedback quantization on the throughput of a multiuser diversity scheme. In *Global Communications Conference* (*GLOBECOM'03*), volume 1. IEEE, December 2003.
- [45] Y. Gao, L. Chen, X. Zhang, and Y. Jiang. Performance evaluation of mobile wimax with dynamic overhead. In *Vehicular Technology Conference*, VTC2008-Fall. IEEE, September 2008.
- [46] D. Gesbert and S. Alouini. How much feedback is multi-user diversity really worth? In International Conference on Communications (ICC). IEEE, June 2004.
- [47] David Gesbert and M.S. Alouini. Selective multi-user diversity. In IEEE International Symposium on Information Theory. IEEE, 2003.
- [48] A. Goldsmith and S. G. Chung. Variable-rate variable-power m-qam for fading channels. *IEEE Transactions on Communications*, 45(10):1218–1230, 1997.
- [49] A. Goldsmith and P. Varaiya. Capacity of fading channels with channel side information. IEEE Transactions on Information Theory, 43(6):1986–1992, 1997.
- [50] I.S. Gradshteyn and I.M. Ryzhik. Table of Integrals, Series, and Products, Seventh Edition. Academic Press, 7th edition, 2007.
- [51] J. Gross, H.F. Geerdes, H. Karl, and A. Wolisz. Performance analysis of dynamic ofdma systems with inband signaling. *IEEE Journal on Selected Areas in Communications*, 24(3):427–436, 2006.

- [52] Myeon gyun Cho, Youngju Kim, Jaekon Lee, and Daesik Hong. A cqi feedback reduction scheme for multi-user mimo/ofdma systems. In *IEEE International Symposium* on Personal, Indoor and Mobile Radio Communications (PIMRC'07). IEEE, September 2007.
- [53] L. Hanzo, C. Wong, and M. Yee. Adaptive Wireless Transceivers. John Wiley & Sons, 2002.
- [54] V. Hassel, M.-S. Alouini, D. Gesbert, and G.E. Oien. Exploiting multiuser diversity using multiple feedback thresholds. In *Vehicular Technology Conference*, VTC2005-Spring. IEEE, June 2005.
- [55] V. Hassel, M.-S. Alouini, G. E. Øien, and D. Gesbert. Rate-optimal multiuser scheduling with reduced feedback load and analysis of delay effects. *EURASIP Wireless Communication Networks*, pages 1–7, 2006.
- [56] J. F. Hayes. Adaptive feedback communications. *IEEE Transactions on Communcation Technology*, 16(1):29–34, 1968.
- [57] S. Haykin. Adaptive Filter Theory. Prentice-Hall, 3rd edition, 1996.
- [58] S. Hu and A. Duel-Hallen. Combined adaptive modulation and transmitter diversity using long range prediction for flat fading mobile radio channels. In *IEEE Globecom*. IEEE, November 2001.
- [59] S. Hu, A. Duel-Hallen, and H. Hallen. Long range prediction makes adaptive modulation feasible for realistic mobile radio channels. In 34th Annual Conference on Information Sciences and Systems, March 2000.
- [60] Tian Hui, Sun Qiaoyun, Dong Kun, and Li Xingmin. A hybrid cqi feedback scheme for 4g wireless systems. In *IEEE Vehicular Technology Conference (VTC'09-Spring)*. IEEE, April 2009.
- [61] J.K. Hwang and J.H. Winters. Sinusoidal modeling and prediction of fast fading processes. In *IEEE Globecom*. IEEE, November 1998.
- [62] Seung-Hoon Hwang, Bonghoe Kim, and Young-Sam Kim. A hybrid arq scheme with power ramping. In *IEEE Vehicular Technology Conference (VTC2001-Spring)*, March 2001.

- [63] Brian L. Evans Ian c. Wong. Optimal resource allocation in the ofdma downlink, with imperfect channel knowledge. *IEEE Transsaction on Communications*, 57(1):232–241, January 2009.
- [64] Satoru Iijima, Rui Zhou, and Iwao Sasase. Reduction of feedback overload by exploiting adaptive channel in ofdma systems. In *Pacific Rim Conference on Communications*, *Computers and Signal Processing*. IEEE, August 2007.
- [65] P. Jain. On the impact of channel quality estimation on adaptive modulation. Master's thesis, Virginia Polytechnic Institute and State University, December 2002.
- [66] M. Johansson. Benefits of multiuser diversity with limited feedback. In 4th Workshop on Signal Processing Advances in Wireless Communications (SPAWC). IEEE, June 2003.
- [67] M. Jordan, N. Hadaschik, G. Ascheid, and H. Meyr. Snr prediction for opportunistic beamforming using adaptive filters. In NEWCOM-ACoRN Joint Workshop, September 2006.
- [68] M. Jordan, L. Schmitt, G. Ascheid, and H. Meyr. Prediction of downlink snr for opportunistic beamforming. In Wireless Communications and Networking Conference (WCNC'07). IEEE, March 2007.
- [69] Eduard Jorswieck, Aydin Sezgin, Bjorn Ottersten, and Arogyaswami Paulraj. Feedback reduction in uplink mimo ofdm systems by chunk optimization. EURASIP Journal on Advances in Signal Processing, 2008.
- [70] Frank Kelly. Charging and rate control for elastic traffic. European Transactions on Telecommunications, 8(1):33–37, October 1997.
- [71] H. Kim and Y. Han. A proportional fair scheduling for multicarrier transmission systems. *IEEE Communications Letters*, 9(3):210–212, 2005.
- [72] R. Knopp and P.A. Humblet. Information capacity and power control in single-cell multiuser communications. In *IEEE International Conference on Communications (ICC'95)*. IEEE, June 1995.
- [73] Marios Kountouris, Thomas Salzer, and David Gesbert. Scheduling for multiuser mimo downlink channels with ranking-based feedback. EURASIP Journal on Advances in Signal Processing, 2008.
- [74] I. Z. Kovács, K. I. Pedersen, T. E. Kolding, A. Pokhariyal, and M. Kuusela. Effects of non-ideal channel feedback on dual-stream mimo-ofdma system performance. In *IEEE Vehicular Technology Conference (VTC-2007 Fall)*. IEEE, 2007.
- [75] Jung Hyoung Kwon, Duho Rhee, Il Mu Byun, Younghoon Whang, and Kwang Soon Kim. Adaptive modulation technique with partial cqi for multiuser ofdma systems. In International Conference Advanced Communication Technology (ICACT'06). IEEE, February 2006.
- [76] Woongsup Lee and Dong-Ho Cho. Cqi feedback reduction based on spatial correlation in ofdma system. In Vehicular Technology Conference (VTC'08-Fall). IEEE, September 2008.
- [77] M. Lei and A. Duel-Hallen. Long range channel prediction and adaptive transmission for frequency hopping communications. In 41st Annual Allerton Conference on Communications, October 2003.
- [78] Jouko Leinonen, Jyri Hamalainen, and Markku Juntti. Performance analysis of downlink ofdma resource allocation with limited feedback. *IEEE Transaction on Wireless Communications*, 8(6):2927–2937, June 2009.
- [79] Qingwen Liu, Xin Wang, and Georgios B. Giannakis. Cross-layer scheduling algorithms with qos support in wireless networks. *IEEE Transactions on Vehicular Technology*, 55(3):839–847, May 2006.
- [80] Qingwen Liu, Shengli Zhou, and Georgios B. Giannakis. Cross-layer combining of adaptive modulation and coding with truncated arq over wireless links. *IEEE Transactions* on Wireless Communications, 3(5):1746–1755, September 2004.
- [81] D.J. Love, R.W. Heath, V. Lau, D. Gesbert, B. Rao, and M. Andrews. An overview of limited feedback in wireless communication systems. *IEEE Journal on Selected Areas in Communications*, 26(8):1341–1365, 2008.
- [82] Informa UK Ltd. Mobile internet traffic: Analysing global usage trends. http://media2.telecoms.com/downloads/mobile-internet-traffic-trends.pdf, 2010.
- [83] MATLAB. version 7.5 (R2007b). The MathWorks Inc., Natick, Massachusetts, 2007.
- [84] Richard Van Nee and Ramjee Prasad. OFDM for Wireless Multimedia Communications. Artech House, 2000.

- [85] T.D. Nguyen and Y. Han. A proportional fairness algorithm wit qos provision in downlink ofdma systems. *IEEE Communications Letters*, 10(11):760–762, 2007.
- [86] Hyeju Oh and Hyung-Myung Kim. An adaptive determination of channel information feedback period in ofdma systems. In 10th International Symposium on Consumer Electronics (ISCE). IEEE, 2006.
- [87] G. E. Oien, H. Holm, and K. J. Hole. Impact of channel prediction on adaptive coded modulation performance in rayleigh fading. *IEEE Transactions on Vehicular Technolo*gies, 53(3):758769, May 2004.
- [88] IEEE P802.11n/D3.00. Part 11: Wireless lan medium access control (mac) and physical layer (phy) specifications: Amendment 4: Enhancements for higher throughput., 2007.
- [89] IEEE P802.16e-2005. Part 16: Air interface for fixed and mobile broadband wireless access systems: Amendment for physical and medium access control layers for combined fixed and mobile operation in licensed bands, December 2005.
- [90] IEEE P802.16m-SSD-2009. Ieee 802.16m system description document (sdd), draft, 2009-09-24, ieee 802.16m-09/0034r2. wirelessman.org/tgm/, September 2009.
- [91] K.I. Pedersen, G. Monghal, I.Z. Kovacs, T.E. Kolding, A. Pokhariyal, F. Frederiksen, and P. Mogensen. Frequency domain scheduling for ofdma with limited and noisy channel feedback. In *IEEE Vehicular Technology Conference (VTC'07-Fall)*. IEEE, October 2007.
- [92] John G. Proakis. Digital Communications. McGraw-Hill, 2001.
- [93] Motorola R1-051334. Cqi feedback scheme for eutra, 2005.
- [94] T.S. Rappaport. Wireless Communications: Principles and Practice. Prentice Hall publications, 1996.
- [95] Adeel Razi and Jinhong Yuan. Feedback reduction schemes for mimo broadcast channels. In IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC'08). IEEE, September 2008.
- [96] S. Sanayei and A. Nosratinia. Exploiting multiuser diversity with only 1-bit feedback. In Wireless Communications and Networking Conference (WCNC'05). IEEE, March 2005.
- [97] D. Schafhuber. Wireless OFDM systems: Channel Prediction and System Capacity. PhD thesis, Vienna Technical University, Vienaa, March 2003.

- [98] David W. Scott. Maltivariate Density Estimation: Theory, Practice and Visualization. Wiley-Interscience, 1992.
- [99] Dieter Shcafhuber and Gerald Matz. Mmse and adaptive prediction of time-varying channels for ofdm systems. *IEEE Transactions on Wireless Communications*, 4(2):593– 602, March 2005.
- [100] Hooman Shirani-Mehr, Daniel N. Liu, and Giuseppe Caire. Channel state prediction, feedback and scheduling for a multiuser mimo-ofdm downlink. In Asilomar Conference on Signals, Systems, and Computers. IEEE, October 2008.
- [101] Chakchai So-In, Raj Jain, and Abdel-Karim Tamimi. Capacity evaluation for ieee 802.16e mobile wimax. Journal of Computer Systems, Networks, and Communications, 1(1):1– 10, April 2010.
- [102] O. Somekh, O. Simeone, A.M. Haimovich, and Y. Bar-Ness. Sum-rate analysis of downlink channels with 1-bit feedback. *IEEE Communications Letters*, 11(2):137–139, 2007.
- [103] Z. Song, K. Zhang, and Y. Guan. Statistical adaptive modulation for qam-ofdm systems. In *IEEE GLOBECOM*. IEEE, November 2002.
- [104] Stelios Stefanatos and Nikos Dimitriou. Downlink ofdma resource allocation under partial channel state information. In *IEEE International Conference on Communications* (*ICC'09*). IEEE, June 2009.
- [105] Mikael Sternad, Tommy Svensson, and Sorour Falahati. Towards systems beyond 3g based on adaptive ofdma transmission. *IEEE*, 95(12):2432–2455, September 2007.
- [106] P. Svedman, S.K. Wilson, L.J. Cimini, and B. Ottersten. Opportunistic beamforming and scheduling for ofdma systems. *IEEE Transactions on Communications*, 55(5):941–952, 2007.
- [107] Taiwen Tang and Jr. Robert W. Heath. Opportunistic feedback for downlink multiuser diversity. *IEEE Communications Letters*, 9(10):948–950, October 2005.
- [108] 3GPP TR-25.913. Requirements for further advancements for evolved universal terrestrial radio access (e-utra) (lte-advanced). Technical Report 25.913, European Telecommunications Standards Institute (ETSI), 2009.

- [109] Matteo Trivellato, Stefano Tomasin, and Nevio Benvenuto. Channel quantization and feedback optimization in multiuser mimo-ofdm downlink systems. In *IEEE Global Telecommunications Conference (GLOBECOM'08)*. IEEE, November 2008.
- [110] David Tse and P. Viswanath. Fundamentals of Wireless Communication. Cambridge University Press, 2005.
- [111] Rui Wang and V.K.N. Lau. Robust optimal cross-layer designs for tdd-ofdma systems with imperfect csit and unknown interference: State-space approach based on 1-bit ack-/nak feedbacks. *IEEE Transactions on Communications*, 56(5):754–761, May 2008.
- [112] Wang Lilei Wang Xiaoyi. A new mathematical model for analyzing cqi channel allocation mechanism in ieee 802.16 systems. In *IEEE International Conference on Circuits and* Systems for Communications (ICCSC'08). IEEE, May 2008.
- [113] W. T. Webb and R. Steele. Variable rate qam for mobile radio. *IEEE Transactions on Communications*, 43(7):2223–2230, 1995.
- [114] Erik Westman. Calibration and evaluation of the exponential effective sinr mapping (eesm) in 802.16. Master's thesis, KTH - Royal Institute of Technology, September 2006.
- [115] C. Wong, R. Cheng, K. Letaief, and R. Murch. Multiuser ofdm with adaptive subcarrier, bit, and power allocation. *IEEE Journal on Selected Areas in Communications*, 17(10):1747–1758, 1999.
- [116] P. Xia, S. Zhou, and G. B. Giannakis. Adaptive mimo ofdm based on partial channel state information. *IEEE Transactions on Signal Processing*, 52(1):202213, January 2004.
- [117] Y. Xue and T. Kaiser. Exploiting multiuser diversity with imperfect one-bit channel state feedback. *IEEE Transactions on Vehicular Technology*, 56(1):183–193, 2007.
- [118] Lin Yang, Mohamed-Slim Alouini, and David Gesbert. Further results on selective multiuser diversity. In International Symposium on Modeling Analysis and Simulation of Wireless and Mobile Systems (MSWiM'04). ACM, October 2004.
- [119] T. S. Yang, A. Duel-Hallen, and H. Hallen. Long range fading prediction to enable adaptive transmission at another carrier. In *IEEE Workshop on Signal Processing Advances* in Wireless Communications (SPAWC). IEEE, June 2003.

- [120] Tung-Sheng Yang, A. Duel-Hallen, and H. Hallen. Reliable adaptive modulation aided by observations of another fading channel. *IEEE Transactions on Communications*, 52(4):605–611, April 2004.
- [121] Ran Yaniv, Danny Stopler, Tal Kaitz, and Kfir Blum. Cinr measurements using the eesm method, April 2005.
- [122] S. Ye, R. Blum, and L. Cimini. Adaptive modulation for variable rate ofdm systems with imperfect channel information. In *IEEE VTC'02*. IEEE, May 2002.
- [123] S. Yoon, C. Suh, Y. Cho, and D. Park. Orthogonal frequency division multiple access with an aggregated sub-channel structure and statistical channel quality measurement. In Vehicular Technology Conference (VTC) 2004 Fall. IEEE, September 2004.
- [124] Randa Zakhour and David Gesbert. A two-stage approach to feedback design in multiuser mimo channels with limited channel state information. In *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC'07).* IEEE, September 2007.
- [125] Erlin Zeng, Shihua Zhu, Xuewen Liao, and Zhimeng Zhong. Impact of limited feedback on the performance of mimo macrodiversity transmission. In *IEEE Wireless Communication* and Networking Conference (WCNC'08). IEEE, March 2008.