UNIVERSITY OF CALIFORNIA, SAN DIEGO

Multiuser diversity and fair resource allocation in wireless heterogeneous networks

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

in

Electrical Engineering
(Communication Theory and Systems)

by

Anh H. Nguyen

Committee in charge:

Professor Bhaskar Rao, Chair
Professor Massimo Francechetti
Professor Philip Gill
Professor William Hodgkiss
Professor Larry Milstein

2014

The dissertation of Anh H. Nguyen is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

____

____

____

____

____

Chair

University of California, San Diego

2014

DEDICATION

To my family.

EPIGRAPH

The price of success is hard work, dedication to the job at hand, and the determination that whether we win or lose, we have applied the best of ourselves to the task at hand.

Vince Lombardi.

A table, a chair, a bowl of fruit and a violin; what else does a man need to be happy?

Albert Einstein.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to express my deep gratitude to my research advisor, Professor Bhaskar Rao, for his patient guidance, enthusiastic encouragement and insightful critiques throughout this dissertation. His guidance and confidence sustained me through times of inspiration and times of difficulty. From him, I gained not only valuable expertise in the field, but also a passion for research. I also wish to thank the members of my PhD committee, Professor Massimo Franceschetti, Professor William Hodgkiss, Professor David Gill and Professor Larry Milstein, for the invaluable knowledge I gleaned from their classes. I am grateful for the time they spent tracking my progress, and for the feedback and direction they provided for my research.

I would like to thank Dr. Zhongren Cao and Joshua Ng for imparting their wisdom and their technical skills. I would like to thank my labmate Dr. Yichao for his help and collaboration for the works in chapter 3, 5, and 6, which proved essential to the project. I also would like to thank my labmate Dr. Eddy for his technical discussions and for sharing his professional experience.

I am grateful to my former and current labmates: Bang Nguyen, Furkan Can Kovasoglu, Sheu Sheu Tan, Yoganada Isukapali, Mathew Pugh, Sagnik Ghosh, and Nandan Das. Over the course of this project they became very good friends, and taught me a great deal about science and about life. I also thank Tu Nguyen, Dung Vo, Son Pham, Tan Nguyen, Cuong Tran, Hieu Nguyen, and Diep Nguyen and other Vietnamese friends for their faithful support.

I would like to thank the Vietnam Education Foundation (VEF) for their fellowship which brought me to UCSD and enabled my graduate study. I could not have completed this project without their financial and moral support. I also wish to express my thank to the Center for Wireless Communications and the National Science Foundation (NSF) grant No. CCF-1115645 for the financial support for my study.

Finally, I would like to thank my parents and my sister for their unflagging support and encouragement. They have been by my side throughout all the ups and downs of this project, and have always pushed me to keep moving forward.

Without them, I could not have arrived at todays accomplishment.

This dissertation is a collection of prior works in which I collaborate with other authors. Chapter 2, in full, is a reprint of the paper User selection schemes for maximizing throughput of Multiuser MIMO systems using Zero Forcing Beamforming, IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2011, co-authored by Anh H. Nguyen and Bhaskar D. Rao. Chapter 3, in full, is a reprint of the paper, Anh H. Nguyen, Yichao Huang and Bhaskar D. Rao, Learning methods used for CDF scheduling in multiuser heterogeneous systems, which is in revision for publication on IEEE Transaction on Signal Processing. Chapter 4, in full, is a reprint of the manuscript, Anh H. Nguyen and Bhaskar D. Rao, "CDF scheduling algorithms for finite rate multiuser systems with limited feedback", which is submitted to IEEE Transaction on Wireless Communications. Chapter 5, in full, is a reprint of the manuscript, Anh H. Nguyen, Yichao Huang and Bhaskar D. Rao, Weighted CDF based scheduling for OFDMA relay downlink with partial feedback, which is preparing to submit to IEEE Transaction on Communications. Chapter 6 is, in full, a reprint of the paper, Anh H. Nguyen, Yichao Huang and Bhaskar D. Rao, Novel partial feedback schemes and their evaluation in an OFDMA system with CDF based scheduling, Asilomar 2013. The dissertation author is a primary author and the other co-authors have contributed or supervised the research for these works.

VITA

2003	BS. in Electronics and Telecommunications, Hanoi University of Science and Technology (HUST).
2005	MS. in Electronics and Telecommunications, HUST.
2014	PhD. in Electrical Engineering (Communication Theory and Systems), University of California, San Diego.

PUBLICATIONS

Anh H. Nguyen and Bhaskar D. Rao, User selection schemes for maximizing throughput of Multiuser MIMO systems using Zero Forcing Beamforming, ICASSP 2011.

Anh H. Nguyen, Yichao Huang and Bhaskar D. Rao, Weighted CDF based scheduling for OFDMA downlink with partial feedback, Asilomar 2012.

Anh H. Nguyen, Yichao Huang and Bhaskar D. Rao, Optimized Quantized feedback in a Multiuser system employing CDF based scheduling, VTC 2013.

Anh H. Nguyen, Yichao Huang and Bhaskar D. Rao, Weighted CDF based scheduling for OFDMA relay downlink with partial feedback, submitted to IEEE Trans. on Communications.

Anh H. Nguyen, Yichao Huang and Bhaskar D. Rao, Performance of a multiuser downlink system applying thresholding feedback with imperfect channel information, VTC 2013.

Anh H. Nguyen, Yichao Huang and Bhaskar D. Rao, Novel partial feedback schemes and their evaluation in an OFDMA system with CDF based scheduling, Asilomar 2013.

Anh H. Nguyen, Yichao Huang and Bhaskar D. Rao, Learning methods used for CDF scheduling in multiuser heterogeneous systems, in revision, IEEE Trans. on Signal Processing.

Anh H. Nguyen, Yichao Huang and Bhaskar D. Rao, "Order statistics based CDF scheduling methods in multiuser heterogeneous systems", ICASSP 2014.

Anh H. Nguyen and Bhaskar D. Rao, "CDF scheduling algorithms for finite rate multiuser systems with limited feedback", submitted to IEEE Trans. on Wireless Commun.

In preparation:

Anh H. Nguyen and Bhaskar D. Rao, "CDF based scheduling algorithms for correlated channel in multiuser systems".

Anh H. Nguyen, Yichao Huang and Bhaskar D. Rao, Feedback design for OFDMA system with CDF based scheduling.

Anh H. Nguyen and Bhaskar D. Rao, "Scheduling, QoS provision and user admission in heterogeneous systems".

ABSTRACT OF THE DISSERTATION

Multiuser diversity and fair resource allocation in wireless heterogeneous networks

by

Anh H. Nguyen

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

University of California, San Diego, 2014

Professor Bhaskar Rao, Chair

In wireless communications, it is of utmost importance to exploit multi-user diversity and at the same time provide satisfactory quality of service for all users. However, these two goals often conflict with each other. On one hand, multiuser diversity is maximized by selecting the user with the best channel condition. On the other, ensuring fairness among users demands the allocation of network resources to those who do not necessarily have the best channel conditions. Whenever a user with a poorer channel condition is selected, there is a certain loss in the overall system throughput. The major objective of this thesis is to find scheduling algorithms that guarantee fairness with minimal performance tradeoff.

First, we consider multi-user diversity in a multi-user MIMO system. When

zero-forcing beam-forming transmission technique is used, the system needs to find a subset of users such that the transmission to these users results in the highest throughput. As the number of users grows, the complexity of the user subset selection increases exponentially. To address this issue, simple user-subset-selection algorithms have been developed that can perform well and are very close to the optimal ones found through an exhaustive search.

Maximizing system throughput is a key factor in ensuring high network performance, but guaranteeing service provision to all users is no less important. To support fairness among users, cumulative distribution function (CDF) scheduling is utilized because of the its capability to precisely control allocation for each user. The CDF scheduling algorithm requires knowledge of the channel distribution among all users. However, the channel distribution or even an approximation of it is hard to obtain in real systems. In this dissertation, two classes of practical, CDF-based scheduling algorithms are developed. They are the non-parametric CDF scheduling (NPCS), used when the channel model is unknown, and the parametric CDF scheduling (PCS), used when the channel model is known. These algorithms are shown to frequently outperform the well-known Proportional Fair (PF) scheduling method, and may be viable alternatives to it. The performance of the developed scheduling technique is then carefully analyzed and verified through simulations under various channel models. In order to apply them in real systems, these algorithms are first proposed for continuous rate transmission. Modified versions are then developed for finite rate transmission and limited feedback resources.

Lastly, we analyze throughput of heterogeneous relay OFDMA systems using CDF scheduling with partial feedback. The scheduling problem is even more challenging with the incorporation of relays because of the different coherent time on their two hops. The CDF scheduling algorithm is modified to satisfy short-term fairness among users. In addition, performance of different feedback schemes in a wideband multi-user system are compared. Among the considered schemes, thresholding feedback is numerically shown to have the lowest feedback requirement, given a certain probability of feedback availability.

Chapter 1

Introduction

In wireless communications, it is critical to exploit multi-user diversity to better utilize network resources and improve the overall system performance. In general, mobile users scattering around the network are very diverse both in channel statistics and in service requirements. There has been much work in the past addressing the problem of multi-user diversity and user service requirements. One extreme solution is maximum rate scheduling, [2] which selects the user with the best channel quality. In maximum rate scheduling, there is no mechanism to make sure every user will be allocated. Because of this, some users who are far from the BS will almost never be allocated resources. Another extreme is round robin as in [3], which sequentially allocates resources to users. This scheduling method supports user fairness because each user can be served equally or proportionally to their requirements. However, round robin fails to exploit multi-user diversity because it does not select users based on their channel quality. As a consequence, maximum rate scheduling has the highest system throughput without guaranteeing service for all users, round robin ensures fairness to users with poor system performance. Because both fairness and performance are key factors of a wireless network, finding a suitable tradeoff between these two priorities is the major goal of scheduling algorithms.

Many scheduling methods have been developed to tackle the problem of balancing performance and fairness. These methods can be classified into temporal fairness [4], proportional fairness (PF) [5, 6, 7], and utilitarian fairness [8]. In temporal fairness, each user is ensured a certain share of time by adding suitable offsets, so that users with larger channel attenuation can still compete for resources. This method requires exact channel distribution in order to precisely control the allocation of resources to users. Its major limitation is that in general, it has no sign of the mention tradeoff optimality. In fact, significant loss in system throughput have been observed [9]. In utilitarian fairness, the scheduler allocates resources in a way that maximizes a selected utility function. The utility function is selected upon the systems preference, and scheduling algorithms often perform well if they are well characterized by the corresponding utility functions. However, when success is measured by fairness and throughput, an algorithm that is optimal in terms of a specific utility function may not necessarily be a good algorithm.

To provide a good tradeoff between fairness and the consumed network resources that are reflected in system performance, PF is proposed in [5]. The PF algorithm considers both the instantaneous rate, which is a reflection of multi-user diversity, and user fairness, which is represented by the amount of resources each user has received in the previous period. The algorithm selects the user with the highest relative fraction between the instantaneous rate and the average rate. PF is a simple and very efficient scheduling algorithm. It can adjust portions of resources for users by tuning its parameters. It can also change the window size used in calculating the previous average throughput, and in doing so, incorporate options in maintaining short-term fairness. However, it is hard to verify whether the tradeoff between fairness and throughput in PF is optimal or not. There might be better algorithms, which suggests the need for further exploration.

The term ”fairness” is very general, and can be understood in a variety of different ways. It might mean that each user gets a similar amount of resources, or it might be interpreted differently. The first definition of fairness does not take into account the diversity in different users service requirements. For example, a user watching real-time video should receive higher throughput than a user working with a text application. This dissertation understand fairness as the allocation of resources to each user according to their needs. The study stresses the prioritization of users. The study also takes into account the fact the systems ability to provide service is location-dependent. We aim to develop scheduling mechanisms that better address users location-dependent characteristics, exploit multi-user diversity, and sufficiently meet users service requirements.

The main focus of my thesis is on cumulative distribution function (CDF) scheduling, which determines users service requirements and prioritizes them based on location-dependent channel characteristics and the distribution functions of the channels. One of the advantages of CDF scheduling is the algorithms ability to precisely allocate predetermined portions of network resources to all users. A major drawback of this algorithm is that it is based on the users channel distribution functions, which are difficult to capture accurately in a real-world context. One of the goals of this study is to develop effective algorithms that can be applied to real networks.

It is observed that spatial and multi-user diversity are very useful in combating fading and in achieving robust communication to the users. By selecting suitable users for transmission, we are able to communicate on good channels and avoid weak ones. The problem of choosing the right user on each spatial mode while ensuring spatial diversity is complex, and this complexity grows exponentially with the size of the system. Because an exhaustive search is not always possible, different user selection algorithms with varying tradeoff between performance and complexity are developed to meet system requirements and computational capability. Among these algorithms, the Greedy approach is the simplest. It sequentially selects a user for each spatial mode in a way that the combination of the newly selected user and the already selected ones results in the highest system throughput increment. Then, the committee machine and evolutionary approaches are proposed. These two algorithms achieve better system throughput and can approach the system throughput using the exhaustive search.

The study also considers fairness and the priority of users. It applies CDF-based scheduling to support priorities of users. A notable characteristic of CDF scheduling is that every user is guaranteed to get a precise, predefined portion of the total network resources. However, there is the fact that channel quality to a user under fast fading can have any distribution. The channel characteristics depend on the transmission environment, which encompasses rich or few scatterers, line-of-sight or non-line-of-sight transmission, the material of the surrounding objects, and many more factors. Because of these factors, a users channel quality can have any distribution which is not restricted, or looks similar to any single constructed model such as Rayleigh, Nakagami-m or log-normal. It is therefore difficult to find a channel model that can represent the distribution of channel gain for any user. Instead, it would be beneficial to develop methods to directly make scheduling decisions from the fed-back CQI without explicitly learning the channel model.

Following this track, the study develops several classes of practical CDF-scheduling algorithms named NPCS and PCS. Interestingly, CDF scheduling can perform efficiently based on the explicit fed-back CQI, even with no knowledge of channel distribution. In real networks with constrained feedback capability, explicit CQI is not generally available. Instead, the base station (BS) only has a quantized version of the CQI or partial feedback. Because of this, it is necessary to modify CDF scheduling to make it suitable for a limited feedback system. Since receiving feedback information requires network resources, it is critical to wisely use the constrained feedback resource to maximize the system throughput. It is also important to find a method of achieving the desired system performance with the minimum feedback requirement.

The study also investigates the application of CDF scheduling to heterogeneous wireless systems. Using CDF scheduling for different devices seems straightforward if we just plug in the channel distribution functions of different users to the scheduler. In fact, it is not that easy. The incorporation of small cells, specifically relays, into a bigger cell makes the scheduling problem a major challenge. The BS must manage the resources allocated for the small cells, and then assign a priority to each small cell based on the total priorities of the users it serves. The channel distribution function of a user served by a relay is now a combination of both the channel from the BS to the relay, and from the relay to that user. Because these two links often have different dynamics, the resource allocation algorithm should be adjusted accordingly in order to suitably provide services to the user.

In summary, the contributions of this thesis are as follows

1.1 Contributions

1.1.1 User selection schemes for maximizing throughput of multiuser MIMO systems using Zeroforcing Beamforming

The performance of a multi-user MIMO broadcast system depends on how the users being served are selected from the pool of users requesting service. Though dirty-paper coding can ensure optimal capacity for multi-user MIMO broadcast systems, it also has high computational complexity. We examine this problem in the context of systems using zero-forcing beamforming (ZFBF), which has low complexity and reasonable throughput. The channels of the selected users may not be orthogonal to each other, which results in harmful interference and reduction in system performance. The study identifies and thoroughly investigates the main factors affecting performance and proposes some mitigating schemes to fully exploit multi-user diversity. In particular, we develop improved user-selection schemes that maximize the sum rate and offer a range of complexity-to-performance tradeoffs. Simulation results show that the selection schemes developed compare favorably with existing techniques.

1.1.2 Learning Methods for CDF Scheduling in Multiuser Heterogeneous Systems

In modern heterogeneous wireless networks, it is both desirable and challenging to support fairness and user priorities, all while achieving the highest possible system. This study develops two classes of practical CDF-based scheduling algorithms to achieve these goals. These algorithms are shown to frequently outperform the well-known Proportional Fair (PF) scheduling method, and may be viable alternatives to PF. The first class of algorithms, Non-parametric CDF-based scheduling (NPCS) algorithms, are used when the channel fading model is unknown. The mapping from channel quality information (CQI) to the real CDF is unknown but is constructed by exploiting the order statistics of the CQI sequence. The constructed CDF mapping methods are shown to converge to the actual CDF. When the channel model is known, a class of Parametric CDF-based scheduling (PCS) algorithms are developed. These algorithms learn the parameters of the channel statistics for the scheduler to use. In our experiments, this Bayesian learning approach results in better system throughput than the NPCS approach. The study also shows that collecting a moderate number of CQI data is enough to achieve nearly the same performance as CDF-based scheduling with known channel distribution. Throughout the study, CDF-based scheduling algorithms are supported by simulations, which show that CDF-based algorithms can effectively support not only fairness but also user priorities. The simulations also show that these algorithms often outperform PF in terms of system throughput.

1.1.3 CDF scheduling algorithms for finite rate multiuser systems with limited feedback

Although CDF scheduling is a promising algorithm, the existing methods were developed assuming that the system will have explicit feedback or know the channel distribution. This study develops simple CDF-scheduling algorithms that can work without knowing the channel distribution, and that can sufficiently support discrete rate transmission. The developed algorithms include window-based extended CDF scheduling, modified NPCS-2, and a hybrid scheme. These algorithms exploit the limited feedback resource for their scheduling decisions and transmission rate selections. They adapt to the empirical distribution of each users channel quality information to best exploit multi-user diversity and the available feedback resources. Then, system throughput of the hybrid scheme is analyzed as a function of the number of collected channel samples. Throughout the study, the developed algorithms are shown to converge quickly to the ideal case of knowing channel CDF, and thus achieve the highest throughput for finite rate systems and effectively save transmit power.

1.1.4 Weighted CDF-based Scheduling for an OFDMA Relay Downlink with Partial Feedback

This section analyzes the performance of partial-feedback OFDMA systems, in which users are served by either a macro BS or by a combination of a macro BS and relays. To reduce feedback, each user feeds back only the best M channel quality information (CQI) among the total number of resource blocks. A weighted, CDF-based scheduling algorithm is used to help ensure diverse service requirements (i.e., fairness) among users, and to exploit multi-user diversity. A general analysis is provided of the system with only a macro BS and the system with both a macro BS and relays. The study reveals that the allocation probability for the users can be met by appropriate choice of weights, which are unique and can be found efficiently. Over the course of the study, I observed that when adding relays to the system, the different time scales on the two hops of the relay cause resource starvation for users. To maintain short-term fairness, I developed a modified version of the weighted, CDF-based scheduling. The system was further enhanced by exploiting the imbalance in the signal-to-noise ratio (SNR) on the two hops of a relay to save transmit power.

1.1.5 A unified comparison of partial feedback schemes in an OFDMA system with CDF based scheduling

In an Orthogonal Frequency Division Multiple Access (OFDMA) system, CDF-based scheduling can be applied to prioritize users. In such a wideband multi-user system, it is essential to reduce the feedback requirement, which is typically very large. This paper considers a system in which several adjacent, highly correlated subcarriers are grouped into a resource block. These resource blocks can be further combined to form subbands, each consisting of a certain number of resource blocks. To reduce feedback, the paper considers and compares three feedback schemes: the opportunistic, the best M and the hybrid. It provides analytical results, which are used to set optimal feedback parameters to exploit system resources while still maintaining fairness among users. It also describes the convergence behaviors of the schemes total feedback requirement. Then, the paper verifies the tradeoff between feedback reduction and system performance by conducting a simulation. The experiments show that the optimal hybrid and opportunistic schemes are identical and result in the lowest feedback requirement.

1.2 Dissertation Outline

The outline of my dissertation is as follows.

In Chapter II, methods to exploit multiuser diversity are developed. Because it becomes more complex to search for an optimal user subset as the system size increases, we propose to use Greedy approach to find a sub-optimal user subset. Then, two other approaches - committee machine and evolutionary - can be used to further improve the quality of the user subset selection. It has been numerically verified that these methods can quickly approach the performance of an exhaustive search with much lower complexity.

In Chapter III, learning methods are developed to make CDF scheduling a practical and competitive algorithm in wireless communications. We propose several classes of algorithms based on the assumption of channel knowledge. When the channel model is unknown, we exploit the order statistics of the fed-back CQI sequence to estimate the CDF of the instantaneous CQI. The estimated CDF value can be the mean, the median or any manipulated version of the real one. A randomizing step can be applied to guarantee the precision of resource allocation for users. When the channel model is known, the channel parameters are learned and then used to calculate the CDF. I then analyze the performance of these methods and find them to compare favorably with the existing PF scheduling methods.

In Chapter IV, we develop CDF scheduling algorithms designed to work best for a system with finite rate transmission and limited feedback. Initially, we propose a window-based extended CS and a modified version of NPCS-2 algorithm. These algorithms work well but each had several weaknesses. Interestingly, the advantages of one algorithm were the weaknesses of the other and vice-versa. We propose a combination, and develop a solution that combines the virtues of the two previously proposed algorithms.

In Chapter V, the application of CDF scheduling to multi-user OFDMA systems with partial feedback is investigated. The paper devotes particular attention to analyzing the system throughput in a single cell with only macro users. The analysis is then applied to systems with both macro users and relays. Because each user has their own priorities, we propose an optimization process to suitably set weights for users based on their activity. The difference in coherent time of two hops of each relay is shown to potentially lead to users resource starvation. We propose a modified CDF scheduling algorithm to prevent such starvation.

In Chapter VI, feedback reduction schemes for CDF scheduling are investigated. In this chapter, the thresholding, best M, and hybrid schemes are compared and analyzed. The thresholding and the hybrid scheme are shown to be identical identical, and they result in the lowest system feedback requirement.

Chapter VII draws conclusions.

Chapter 2

User selection schemes for maximizing throughput of multiuser MIMO systems using Zeroforcing Beamforming

The performance of a multiuser MIMO broadcast system depends highly on how the users being served are selected from the pool of users requesting service. Though dirty paper coding can obtain the optimal capacity of multiuser MIMO Broadcast systems, the computational complexity is high. We examine this problem in the context of systems using Zero Forcing Beamforming (ZFBF) which have low complexity and reasonable throughput. The channels of the selected users may not be orthogonal to each other resulting in harmful interference and reduction in the performance of the system. This work identifies and thoroughly investigates the main factors affecting the performance and proposes some mitigating schemes to fully exploit multiuser diversity. In particular, we develop improved user selection schemes designed to maximize the sum rate that offer a range of complexity to performance tradeoff. Simulation results show that the selection schemes developed compare favorably with existing techniques.

2.1 Introduction

Multiuser MIMO has been a topic of interest in recent years and is well known for significantly enhancing system's capacity as it offers both multiuser diversity and spatial diversity [10]. Many algorithms have been developed offering a tradeoff between complexity and performance. In [11], a iterative water filling algorithm is proposed which offers good performance but is complex. The complexity of the iterative algorithm proposed is high due to not only its dependence on the size of the system but also on the high number of iterations required for the algorithm to converge. The duality between the uplink and downlink [12] allows one to have an equivalent result on the downlink using results from the uplink. To reduce complexity, [13, 14, 15] proposed the use of Zero Forcing Beamforming (ZFBF) which can be further combined with interference cancellation in [16].

In a multiuser system, the way users are selected is of utmost importance to fully exploit the multiuser diversity. Finding the best subset of users has combinatorial complexity and is practically infeasible. In [14], a Semi-Orthogonal User Selection (SUS) scheme was proposed. It is a greedy selection technique based on the projection of users' channels on the space orthogonal to the space formed by channels of previously selected users. Though this selection scheme is intuitive and has high performance, there are still questions on its optimality and consequently on whether we can do better. In this paper, we develop user selection schemes that allow for a tradeoff between complexity and performance.

2.2 System Model

We consider a Multiuser MIMO downlink (Broadcast) system in the following discussion. Similar results can be obtained for the uplink system. We assume there are K users with each user equipped with M_R antennas and the base station has M_T antennas. The channel gains of users are assumed to be i.i.d. Rayleigh and the transmitter has full CSI. The received signal of user k, at time instant i, is

$$\mathbf{y}_{k,i} = \mathbf{H}_k \sum_{j=1}^{K} x_{j,i} + n_{k,i} = \mathbf{H}_{k,i} x_i + \mathbf{n}_{k,i}$$

$x_i = \sum_{j=1}^{K} x_{j,i}$ is the sum of signals intended for all K users. A block fading model is assumed and $\mathbf{H}_k$ is $M_R \times M_T$ MIMO channel of user k. $\mathbf{n}_k$ is complex, spatially and temporally white noise $(CN(0, I))$ at user k.

The problem of capacity maximization for a Multiuser MIMO systems consists of maximizing the sum capacity over the design parameters which include power allocation with a total power constraint and the transmit and decode strategies. The optimal resource allocation for this system is complicated and simpler techniques have been developed. Among the techniques developed for solving this problem, techniques based on ZFBF, QR decomposition, and Cooperation between users are widely known for their simplicity and high performance. Depending on the technique used, we have to match the user selection approach to maximize system throughput. In this paper we also consider ZFBF and develop appropriate user selection schemes.

It is assumed that each user uses maximum ratio transmission, i.e. the beam forming vector is $v_{1,k}$, the right singular vector corresponding to the dominant singular value $\sigma_{1,k}$ of $\mathbf{H}_k$, and at the receiver MRC combining is employed using the corresponding left singular vector $u_{1,k}$. The scalar decision variable at receiver k is

$$y_{k,i} = u_{1,k}^H \mathbf{y}_{k,i} = u_{1,k}^H \mathbf{H}_k x_i + u_{1,k}^H \mathbf{n}_{k,i} = \sigma_{1,k} v_{1,k}^H x_i + n_{k,i}$$

Assuming r users are served, for the analysis we stack the decision variables at the different receivers to obtain $y_i = [y_{1,i}, \dots y_{r,i}]^T$. After some manipulation, we have

$$y_i = \Sigma \mathbf{V}^H x_i + n_i, \tag{2.1}$$

where $\Sigma = diag\,[\sigma_{1,1}, \dots, \sigma_{1,r}]$ and is formed from the largest singular values of the r users and $\mathbf{V} = [v_{1,1} \dots v_{1,r}]$ is formed with the right singular vectors of the users' channels. Please note that $\mathbf{V}$ is not normally an unitary matrix.

Before we present our selection schemes, we describe the SUS algorithm [14] which is widely used and serves as a reference for comparison. In SUS, $h_k = \sigma_k v_{1,k}$ is used in the selection process as follows.

- Initialization: mode $i = 1$; $T = \phi$. **Iteration**
- For each user k, calculate g_k, the component of h_k orthogonal to the subspace spanned by $\{g_{(1)}, \dots, g_{(i-1)}\}$.
  - $g_k = \left(I - \sum_{j=1}^{i-1} \frac{g_{(j)} g_{(j)}^H}{\|g_{(j)}\|^2} \right) h_k$. which is the projection of the channel vector onto the space orthogonal to channel vectors of the previously selected users.
- Select the i^{th} user as follows:
  - $k_i = \arg\max_k \|g_k\|$; $T \leftarrow T \bigcup \{k_i\}$; $h_{(i)} = h_{k_i}$; $g_{(i)} = g_{k_i}$

Until $i = r$

T contains the indices of the selected users.

2.2.1 Throughput and optimal user selection scheme for the system using ZFBF

Since $\mathbf{V}$ is not unitary, we perform a QR decomposition, $\mathbf{V} = \mathbf{QR}$, and we can rewrite equation (2.1) for the Broadcast Channel

$$\begin{aligned} y_i &= \Sigma \mathbf{V}^H x_i + n_i \\ &= \Sigma \mathbf{R}^H \mathbf{Q}^H \mathbf{Q} \tilde{x}_i + n_i \\ y_i &= \Sigma \mathbf{R}^H \tilde{x}_i + n_i \end{aligned} \tag{2.2}$$

$x_i = \mathbf{Q}\tilde{x}_i$, is the transmitted signal and $\tilde{x}_i$ is still to be determined and includes the beam forming matrix. To enable decoding the received signals in a decoupled manner, we diagonalize the matrix $\Sigma\mathbf{R}^H$ by the beamforming matrix $\mathbf{W}$ at the transmitter. The sum rate optimization problem then becomes

$$\begin{aligned} \max_W \quad & \log |I + \Sigma \mathbf{R}^H \mathbf{W} \mathbf{W}^H \mathbf{R} \Sigma| \\ s.t. \quad & Tr(\mathbf{W}\mathbf{W}^H) \leq P \end{aligned} \tag{2.3}$$

Normally ZF technique uses a weight matrix which is the inverse of the channel matrix. This not only decouples the users making mutual interference zero, the channel to each user is of equal quality. Here, we use a similar version of the ZF technique where only decoupling is enforced.

$$\begin{aligned} \max_\Lambda \quad & \log |\mathbf{I} + \Sigma \mathbf{R}^H \mathbf{R}^{-H} \Lambda \mathbf{R}^{-1} \mathbf{R} \Sigma| \\ s.t. \quad & Tr(\mathbf{R}^{-H} \Lambda \mathbf{R}^{-1}) = Tr(\mathbf{R}^{-1} \mathbf{R}^{-H} \Lambda) \leq P \end{aligned} \tag{2.4}$$

where Λ is diagonal with $\mathbf{R}^{-H}\Lambda\mathbf{R}^{-1} = \mathbf{W}\mathbf{W}^H$. Because the selected users have good channel gains, the condition number of the equivalent channel is small, and we can assume with adequate total power that every spatial modes is excited. The throughput of the system using ZF technique is determined by

$$C_{ZF} = \sum_{i=1}^{M} \log \left[\frac{1}{r} \left(P + \sum_{l} \frac{(\tilde{\mathbf{R}}^2)_{l,l}}{\sigma_l^2} \right) \frac{\sigma_i^2}{(\tilde{\mathbf{R}}^2)_{i,i}} \right] \tag{2.5}$$

where $\tilde{\mathbf{R}} = \mathbf{R}^{-1}$ and its (i, j)th entry is denoted by $\tilde{\mathbf{R}}_{i,j}$. Since there are K users in the system, selecting a subset of r users in an optimal manner is computationally complex. Thus, we maximize the system throughput by selecting the desired users in a Greedy manner. We select the users sequentially ensuring that at each iteration the new user chosen maximizes the obtainable throughput C_{ZF}, while retaining the users already chosen. This is in contrast to the popular SUS technique that concentrates on user orthogonality and on mutual interference in a an ad-hoc manner. The proposed selection technique requires a sequential update of the QR decomposition of $\mathbf{V}$ which is conducted using a Gram Schmidt process to find $R_{i,i}$, $R_{i,j;j<i}$ (page 224 of [17]). The steps of the selection procedure are as follows.

- Initialize: Selected users $T = \emptyset$, $\mathbf{Q} = \mathbf{I}_r$
- For spatial mode $i = 1$ to r
  - For user $k = 1$ to K
    - Transform $v_{1,k}$ using Q: $v_{1,k} \leftarrow Qv_{1,k}$
    - Using $v_{1,k}(i : r)$, generate Householder matrix $\tilde{\mathbf{Q}}_k$; then $\mathbf{Q}_k = \begin{pmatrix} \mathbf{I}_{i-1} & 0 \\ 0 & \tilde{\mathbf{Q}}_k \end{pmatrix}$
    - Form the i^{th} column of $\mathbf{R}$ by multiplying $\mathbf{Q}_k$ with $v_{1,k}$. Then compute values of $(\tilde{R}^2)_{j,j} = \sum_{l=j}^{i} (\tilde{R}_{j,l})^2$ with $j = 1, \ldots, i$; $\tilde{R}_{j,i}$ is updated recursively as in equation (2.7) [18].
    - Set $\gamma_k = \sum_{j=1}^{i} \log \frac{\sigma_j^2}{(\tilde{\mathbf{R}}^2)_{j,j}}$.
  - Select the user who results in the largest value of γ_k. We denote this user as k_i
  - $\mathbf{R} \leftarrow \mathbf{Q}_{k_i}[\mathbf{R}, v_{1,k_i}]$; $\mathbf{Q} \leftarrow \mathbf{Q}_{k_i}\mathbf{Q}$; $T \leftarrow T \cup k_i$
- End

Though the selection scheme is constructed to be independent of the system's power, we can easily add power to our selection criteria through water level μ and water filling algorithm as in formula 2.5 to further fine tune the selection.

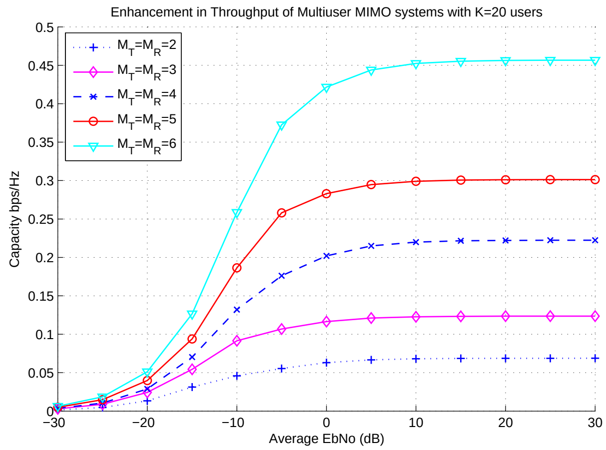


Figure 2.1: Enhancement of the Selection Scheme vs SUS, $K = 20$ users and different antenna configurations (2×2 to 6×6).

In the view of a successive selection process, we can rewrite the throughput of our systems when the i^{th} user is selected as follows:

$$C_{ZF} = i \log \mu + \sum_{j=1}^{i} \log \left(\frac{\sigma_j^2}{(\tilde{R}_{j,j})^2 + \sum_{k=j}^{i} (\tilde{R}_{j,k})^2} \right) \tag{2.6}$$

When a new user is selected, a new mode is added and the channel gains of previous modes are reduced as $(\tilde{\mathbf{R}}^2)_{j,j}$ increases an amount $(\tilde{\mathbf{R}}_{j,i})^2$.

This process is somewhat similar to the SUS in the sense users are successively selected. The metric used to select users is derived optimally for ZFBF in a Greedy manner. Whenever a new user is added, one more spatial mode is exploited and more interference is introduced to the existing spatial modes. While SUS only selects the best spatial mode, our selection method considers the increased interference to the existing modes also. To verify the performance of the proposed selection scheme, we conduct a simulation study. We consider a multiuser MIMO system with $K = 20$ users with different antenna configurations in a flat Rayleigh fading environment. Figure 2.2.1 shows the results. The difference in throughput is plotted and the proposed scheme outperforms SUS. When there are more users in the system, the impact of user selection diminishes slightly.

2.2.2 A simplified selection scheme

Our proposed scheme has many advantages over existing techniques as the used QR decomposition can be computed very efficiently. Furthermore, matrix inversion can be updated in each step and the summation of i logarithms functions at step i is simple also. However, compared with SUS, our selection scheme is more complex. Thus, we further simplify the algorithm so it becomes more suitable for practical systems with a slight reduction in the performance.

Theorem 1. *The off-diagonal elements of* $\tilde{R} = R^{-1}$ *have small values and can be approximated by* $\tilde{R}_{j,i} \simeq R^{-1}_{j,j} R_{j,i} R^{-1}_{i,i}$.

Proof. Assume we are at step i, the element j in column i will be determined as follows [18]:

$$\tilde{R}_{j,i} = \sum_k \tilde{R}_{j,k} R_{i,k} R^{-1}_{i,i} = R^{-1}_{j,j} R_{j,i} R^{-1}_{i,i} + \sum_{k=j+1}^{i-1} \tilde{R}_{j,k} R_{k,i} R^{-1}_{i,i} \tag{2.7}$$

where $R_{i,k}$ is the k^{th} element in column i of the matrix R from the potential selected i^{th} user. $\tilde{R} = R^{-1}$ is the inversion of matrix R at each step. Since the selected user is likely to be orthogonal to the previously selected users, $\tilde{R}_{j,k}$ and $R_{i,k}$ are likely to be small and we can neglect the second term in 2.7 leading to

$$\tilde{R}_{j,i} \simeq R^{-1}_{j,j} R_{j,i} R^{-1}_{i,i}.$$

This enables us to have a much simpler way to update the elements of the inversion matrix. □

As the updates are small, we approximate the change in the cost functions $c_j = \log \frac{\sigma_j^2}{(\tilde{\mathbf{R}}^2)_{j,j}}$ by $\Delta c_j = \frac{\partial c_j}{\partial (\tilde{\mathbf{R}}^2)_{j,j}} \Delta(\tilde{\mathbf{R}}^2)_{j,j}$. Defining $\beta = \log e$,

$$\Delta c_j = -\beta \frac{\Delta(\tilde{\mathbf{R}}^2)_{j,j}}{(\tilde{\mathbf{R}}^2)_{j,j}} \simeq -\beta \frac{(R^{-1}_{j,j} R_{j,i} R^{-1}_{i,i})^2}{R^{-2}_{j,j}} = -\beta \frac{R^2_{j,i}}{R^2_{i,i}}$$

which represents the loss in throughput at spatial mode j. We select a new user to maximize the increment in system throughput

$$\Delta c = \log\left(\sigma_i^2 R_{i,i}^2\right) - \beta \sum_{j=1}^{i-1} \frac{R_{j,i}^2}{R_{i,i}^2} = \log\left(\sigma_i^2 R_{i,i}^2\right) - \beta \left(\frac{1}{R_{i,i}^2} - 1\right)$$

The last step is a consequence of the unit norm singular vectors comprising the matrix $\mathbf{V}$ whose QR decomposition is being computed. The simplified algorithm is performed in a similar manner as the optimal Greedy selection except for the step of calculating $\tilde{R}$ and the change in selection metric. We just select users based on the following metric

$$\gamma_k = \log(\sigma_i^2 R_{i,i}^2) - \beta \frac{1}{R_{i,i}^2} \tag{2.8}$$

Simulations are carried out as before and the simplified version offers considerable improvement over SUS as shown in figure 2.2. The plot shows the difference in throughput on average. This simplified selection scheme requires QR decomposition and calculation of the simple cost function. Thus, it can be considered as simple as SUS with better performance as shown in the simulation. Due to the verified advantages, the simplified algorithm might be a potential replacement of SUS.

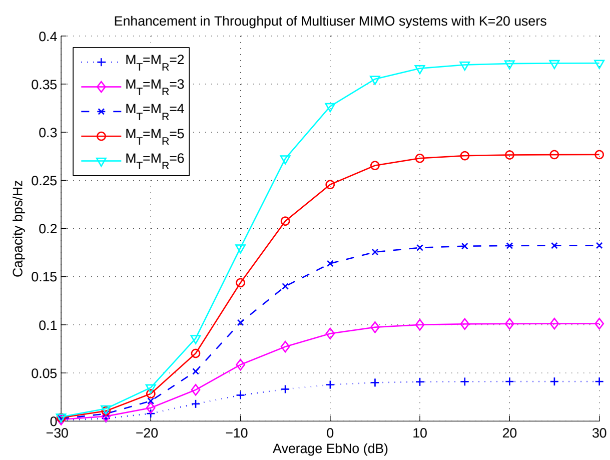


Figure 2.2: Enhancement of the Simplified Scheme vs SUS, $K = 20$ users and different antenna configurations (2×2 to 6×6).

2.2.3 Committee Machine Approach

Even though the above mentioned methods developed perform better than SUS, this is only in the average sense and not always. Furthermore, when the size of the system increases, the enhancement is not very consistent. This suggests combining the proposed scheme with SUS and selecting the better of the two to further improve the performance. The combination with SUS gives us consistent performance with insignificant increase in complexity. Moreover, the two selection schemes can be executed in parallel or in a combined manner with insignificant increase in complexity as the QR decomposition also provides $R_{i,i}$, the projection on the orthogonal space with the space spanned by previously selected beams as a byproduct which can be used in SUS. This line of thinking naturally leads to the committee machine approach [19], wherein there are multiple reasonable low complexity selection schemes and one uses the best from this collection.

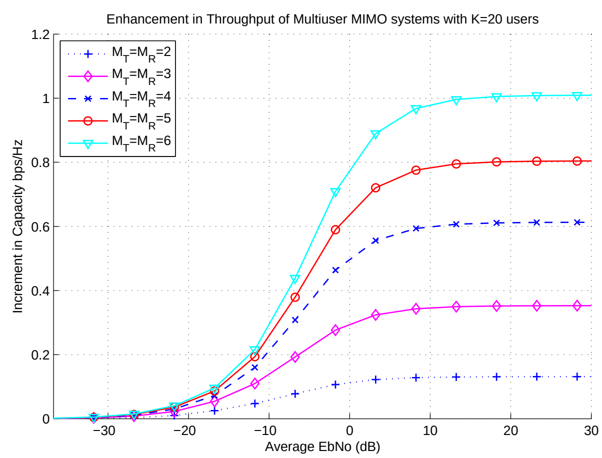


Figure 2.3: Committee machine with 5 Selection Schemes vs SUS, $K = 20$ users and different antenna configurations (2×2 to 6×6).

We now suggest how one may generate several reasonable low complexity selection schemes. Our investigation shows users with larger singular values have higher probability of being in the optimal subset found by the exhaustive search. Thus, we can initially select a user who has a large singular value, not necessarily the largest, and carry out the Greedy selection scheme. The performance is improved drastically as we can search from different initial points and have a higher chance of selecting a better set. Simulations confirm this improvement. In figure 2.3, the system's performance is improved drastically when we use a committee approach, wherein each member procedure is a Greedy approach with a different initial user.

We explore limitations of the approach by comparing the performance of this committee machine with that of an exhaustive search. The results are shown in Figure 2.4. However, there is still a considerable performance gap and we explore a different approach to close this gap, albeit more complex.

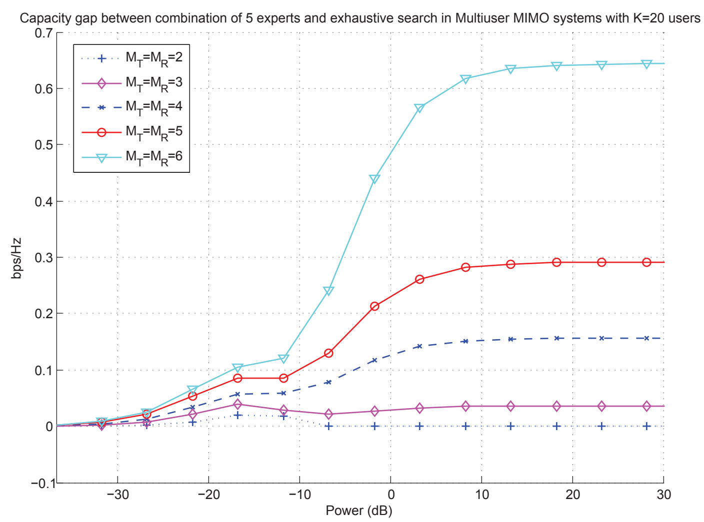


Figure 2.4: Gap between committee machine with 5 searches initiated from 5 users with largest channel gains and exhaustive search, $K = 20$ users and different antenna configurations (2×2 to 6×6).

2.2.4 Evolutionary Approach

Here we use an evolutionary approach to selecting the subset of users. The computation at any given iteration is as follows. At iteration i, we will have l survivors each being a subset of i users. At the $(i + 1)$th iteration, each of the l survivors will have l children leading to a total of l^2 subsets of $(i + 1)$ users. The children can be selected by using the top candidates from the Greedy procedures discussed in the previous sections. The l^2 subsets are then pruned to generate l survivors by selecting the best l. The iterations are repeated till each of the survivors have subset of r users. The best subset is used for ZFBF. This is clearly more complicated but the performance results are quite encouraging. The complexity can be controlled by controlling the branching factor l. The simulation results are shown in figure 2.5. It shows the gap between our selection and exhaustive search can be quickly forced to zero when the number of branches increases. We also note that with smaller size systems, this gap is almost equal to zero.

2.2.5 Complexity analysis

- For downlink systems, we have $r = M_T$ spatial modes
- Each mode, we apply our calculation for K users as follows
  - the complexity of the QR decomposition over all r spatial modes is $O\left(M_T^3\right)$ for each user. At each mode, we have to calculate for K users so the complexity is $O\left(KM_T^3\right)$
  - when multiply Q_kQ with $v_{1,k}$ to form the i column of R, we perform two multiplications of a matrix with a column. This requires $2M_T^2$ multiplications and M_T additions
- r updates of R and Q require $r \times M_T^3 = M_T^4$ multiplication

As the problem is valid only when $K > M_T$, the overall complexity order is $O\left(KM_T^3\right)$ which is similar to that of SUS. The complexity order of the simplified version is still the same but the computational requirement in each step is trivial which can be considered as simple as SUS. Interestingly, our proposed algorithm obtains higher performance with the same computational complexity.

Lastly, the algorithm using the committee approach has superior performance but with additional computational complexity. The computational requirements are increased by a factor of l, where l is the number of selection procedures in the committee. The evolutionary approach also has a complexity comparable to the committee approach with l representing the branching factor.

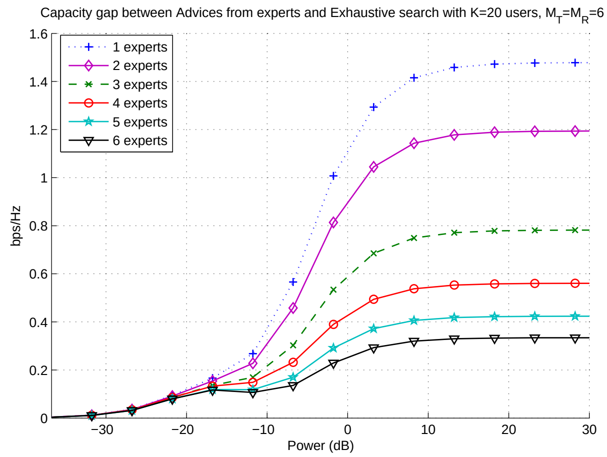


Figure 2.5: Gap between the evolutionary approach, exhaustive search, and SUS $K = 20$ users, $M_T = M_R = 6$ antennas

2.3 Conclusion

We propose user selection schemes with a range of complexity versus performance tradeoff. Our proposed schemes outperform existing methods in either performance or simplicity. Moreover, the idea of combining selection procedures in the committee framework results in significantly improved performance with modest increase in complexity. The evolutionary approach also shows much promise.

The text of this chapter, in full, is a reprint of the paper [20], Anh H. Nguyen, and Bhaskar D. Rao User selection schemes for maximizing throughput of multiuser MIMO systems using Zero Forcing Beamforming, that has been published in IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2011. The dissertation author is the primary researcher and author, and the co-author listed in this publication directed and supervised the research constitutes this chapter.

Chapter 3

Learning methods used for CDF Scheduling in Multiuser Heterogeneous Systems

In modern heterogeneous wireless networks, the task of supporting fairness along with user priorities and concurrently achieving the highest possible system throughput is desirable and challenging. In this work, two classes of practical cumulative distribution function (CDF) based scheduling algorithms are developed to achieve these goals. These algorithms are shown to frequently outperform, and are potential alternatives to, the well known Proportional Fair (PF) scheduling method. The first class of algorithms, Non-parametric CDF based scheduling (NPCS) algorithms, are used when the channel fading model is unknown. Herein, the mapping from channel quality information (CQI) to the real CDF is unknown but is constructed exploiting the order statistics of the CQI sequence. The constructed CDF mapping methods are shown to converge to the actual CDF. When the channel model is known, a class of Parametric CDF based scheduling (PCS) algorithms are developed which learn parameters of the channel statistics for the scheduler to use. In our experiments, this Bayesian learning approach results in better system throughput than the NPCS approach. We also show collecting a moderate number of CQI data is enough to achieve nearly the performance of CDF based scheduling with known channel distribution. Throughout the work, CDF based scheduling algorithms are supported by simulations which show that they can effectively support not only fairness but also user priorities and often outperform PF in terms of system throughput.

3.1 Introduction

In wireless communications, exploiting multiuser diversity is critical to improving the overall system performance [2]. In general, mobile users can be at any location in a wireless network service area which leads to a diversity in pathloss, fading condition and channel statistics. On one hand, to utilize multiuser diversity, the base stations (BS) needs to allocate the right resource to the right user to maximize spectrum usage efficiency. On the other hand, guaranteeing fairness and services to all the users is also of utmost importance in order to ensure that all the users are adequately served. Typically, these objectives conflict and so a compromise between system throughput and fairness has to be made. Among the existing scheduling methods, proportional fair scheduling (PF) [5, 6] is widely used because it has a good balance between performance, fairness and simplicity. However, the resource allocation for users in PF is not easy to control and might need adjustments overtime to satisfy users's requirements in a long term manner. Also, even though it exploits multiuser diversity, the efficacy of the approach is unclear in an heterogeneous environment. This calls for better algorithms and we consider cumulative distribution function (CDF) based scheduling in this context. CDF based scheduling, which is called CS in [21], can control resource allocation precisely and can exploit multiuser diversity effectively in a heterogeneous environment. However, its comparative performance is unknown and the general approach has received much less attention. As a result, there are no or limited practical implementations of CDF based scheduling [1]. A principal challenge in implementation is that the scheduling method requires knowledge of the CDF of each user which is typically unknown, changing frequently and location dependent. Motivated by the favorable properties of CDF scheduling, we develop two classes of algorithms, non-parametric and parametric CDF based scheduling, to make CDF scheduling practical and along the way show its advantages over the widely used PF scheduling.

As stated above, exploiting multiuser diversity is very important. The sum rate capacity of a multiuser system scales up with the number of users due to the increasing multiuser diversity [22]. In [2], the uplink in a single cell multiuser communications is studied and the maximum capacity is achieved considering the effect of multiuser diversity. To achieve the highest system throughput, the BS selects and communicates with the user who has the best channel. To do so, channel information from the users is needed which is available through feedback [23, 24]. As feedback information is beneficial, the area of exploiting channel knowledge has been extensively studied with the major focus on instantaneous channel state information (CSI). The double logarithmic capacity growth of full CSI systems is obtained even with an opportunistic one bit feedback [25]. In the opportunistic approach [22, 25], the users who have poorer channel conditions tend to be ignored.

Because guaranteeing services for all users is of utmost importance in wireless communication systems, various fair scheduling methods have been developed. Several methods are discussed in [26] while reviews of scheduling algorithms in [27, 28] show the developed algorithms can be used for a single cell or multi-cell and aim at different objectives such as capacity, Quality of Service (QoS), fairness, distributed computation, etc. Among fairness algorithms, the round robin (RR) scheduling is known for its high degree of fairness. However, as the algorithm is unable to exploit multiuser diversity, significant loss in system performance is observed [29]. Another algorithm, Nash bargaining, can be applied in OFDMA networks [30]. Meanwhile, the convergence conditions of Nash bargaining and Proportional Fair (PF) are investigated in [31]. Fairness algorithms can be controlled by tuning parameters in the utility function utilized in [32]. They have wide a range of applications such as in a multi-cell context in [33], in OFDM relay networks in [34], or in ad hoc networks in [35]. It is shown in [36] that OFDMA systems employing fairness scheduling can concurrently support good fairness among users and achieve high throughput. From the performance point of view, fairness algorithms showed in [37] can guarantee fairness but at the same time exploit frequency, multiuser and time diversity to enhance overall system throughput.

Though many algorithms have been developed, PF scheduling technique proposed in [5] is widely used because it offers a good tradeoff between exploiting multiuser diversity and maintaining fairness among users. The algorithm is investigated in [38] in a single cell environment and its simplified version is considered in [39]. PF has been used in CDMA systems [40], applied together with antenna selection algorithms in multiuser space division MIMO downlink [41] and applied in LTE [42]. The performance of PF in multiuser OFDMA systems can be optimized, as in [43], given proportional rate constraints are satisfied. The algorithm in the long term is shown to converge to the solution of a reasonable maximization problem in [44]. There are many variations of proportional fair scheduling and their performance are compared with RR in [45]. Variations of PF are also compared from an energy efficiency perspective in [46]. Though being a very effective scheduling technique, under certain condition of the data arrival process the PF is shown to be unstable [47]. In general, PF is hard to analyze and the approach is unable to exactly control resource allocation for users. Furthermore, its effectiveness in a heterogeneous environment also leaves room for improvement.

A viable alternative is the cumulative distribution function (CDF) based scheduling proposed in [21]. It is shown to be able to control precisely the probability of resource allocation for users, exploit multiuser diversity and deal with heterogeneous environments. The CDF-based scheduling is leveraged in a general multicell network in [48] and in a partial feedback OFDMA relay system in [49] to guarantee scheduling fairness and simultaneously obtain multiuser diversity gain. In [50], the CDF-based scheduling is analytically studied in the random beamforming framework, and the notion of individual sum rate and individual scaling laws are proposed to characterize the performance under this scheduling policy. All of the mentioned works assume knowledge of the CDF of the channel, which is an idealistic and simplifying assumption. Taking into account that the CDF is in general unknown and hard to learn [21], a nice practical scheme is proposed in [1] which can be applied to systems which support discrete rates. However there are unanswered questions with respect the optimality of the approach and also the convergence behavior, an important consideration in channel with short coherence time. Despite these concerns, the method developed in [1] is very interesting, motivational, and an important step in making CDF scheduling practical. Quantile scheduling [51] and ranking based scheduling [52] have equivalent scheduling metric and work well when all users have the same priority. Because these algorithms can not guarantee precise resource allocation when users have different priorities, a different approach is therefore needed. Our approach in this paper is to examine CDF scheduling without being constrained by the finite number of discrete data rates and develop effective techniques. The hope is that the methods develop can provide insight, advance the state of the art, and lead to potential alternatives to the discrete data rate case, if desired.

The outline and contributions of this work are as follows.

- In section 6.1, the system model and a brief comparison between CDF based scheduling and PF are shown to motivate the work. Interestingly, the CDF based scheduling is significantly better than RR [1] and can achieve the optimal throughput of Opportunistic scheduling under certain network conditions. This shows CDF based scheduling might frequently compare favorably with PF scheduling.

- In section 3.3, we propose to utilize historical data to capture more detailed information about the distribution, not just the mean as in PF scheduling, for scheduling purposes. In this part, a class of non-parametric CDF based scheduling (NPCS) algorithms are developed. In NPCS, we propose to use order statistics of the CQI sequence of each user and work directly on the CDF of this CQI sequence to help the scheduling process. The NPCS algorithms are used when the channel model such as pathloss, type of fading is not known at the BS. This class of algorithms can be understood as, but not restricted to, exploiting empirical CDF. In particular, they can be cast in a framework to maximize an objective function of choice such as throughput, bit error rate, among others. The variables of optimization are the mapping functions which reflect how the CQI sequence is used for scheduling. Then, two algorithms NPCS-1 and NPCS-2, which aim at best exploiting the ordered CQI are proposed and analyzed. Among all NPCS algorithms, NPCS-2 is proved to be the optimal algorithm in terms of system throughput. When compared to the algorithm in [1], though there are some similarities in the mapping, NPCS-2 has better user selection and can achieve higher system throughput, given the same average transmit power. The enhancement in our experiments is about 10%, which can vary depending on system setup. Another advantage of NPCS-2 is that each user memorizes and orders its own channel CQI and feeds back only the ordered index. This helps to offload some processing requirement to the users.

- In section 3.4, a class of parametric CDF based scheduling (PCS) algorithms is developed. A PCS algorithm is useful when the model of the channel statistics, e.g. the parametric form of the fading is known. The learnt parameters of the channel CDF is then used by the CDF based scheduler. The advantage of this algorithm is that it can achieve higher system throughput with a slight compromise on the ability to precisely control the resource allocation. By collecting a suitably large number of channel CQI, PCS algorithm can achieve arbitrary precision in resource allocation for users.

- In section 3.5, the throughput of the proposed algorithms are compared and shown to be superior to the widely used PF scheduling. Though the throughput of these algorithms are derived in Rayleigh fading, numerical results in Nakagami-m fading are also shown with similar conclusions. It is shown that the algorithms developed approach the performance of a CDF scheduler with perfect knowledge very rapidly. When compared with the performance achieved with exact CDF knowledge, the algorithms achieve $95 \rightarrow 97\%$ of the performance with only 10 i.i.d. channel samples, and achieve 99% the performance with only 30 i.i.d. samples.

3.2 System model and Motivation

3.2.1 System model

We consider a multiuser system where the base station (BS) has a single antenna and the K users are also each equipped with a single antenna. A downlink system is specifically assumed in the presentation and similar considerations apply to the uplink system. At a given time, the BS selects a user k and transmits a symbol s_k to this user. The received signal y_k is

$$y_k = h_k \sqrt{\rho} s_k + n_k, \tag{3.1}$$

where $h_k \in C^{1\times 1}$ is the channel from the BS to the selected user k which is assumed to be independent in time, $n_k \sim CN(0,1)$ is the additive noise at user k and ρ is the transmit SNR. The instantaneous CQI z_k and SNR x_k are given by $z_k = |h_k|^2$ and $x_k = \rho z_k$. Also, we denote the random variables associated with the CQI and SNR of user k by Z_k and X_k respectively. Throughout this work, we use upper case letters, e.g. Z, to denote a random variable and the corresponding lower case letters, e.g. z, to denote a certain value for that random variable.

To select a user to be served on the resource, CDF based scheduling [21, 49] is used. Upon receiving the CQI z_k from users, fed back through an appropriate feedback channel, the BS utilizes the corresponding CDF of these CQI to evaluate a service metric and selects the user k^* with the highest value. The CDF scheduler selects the user to be served as follows:

$$k^* = \arg\max_k F_{Z_k}(z_k)^{\frac{1}{w_k}}, \tag{3.2}$$

where $F_{Z_k}(.)$ is the CDF of the CQI of user k. The weight w_k represents the priority assigned to user k, which is equivalent to the proportion of the resource allocated by the network to the user. The weight w_k is preassigned for all the users $k = 1, \ldots, K$ such that $\sum_{k=1}^{K} w_k = 1$.

Several interesting and desirable attributes of CDF based scheduling have been derived in [21, 49]. We now prove some additional interesting properties of CDF based scheduling to motivate the practical implementation.

3.2.2 CDF based scheduling in comparison with Proportional Fair Scheduling

For this discussion, we assume that the exact CDF of the CQI is known. This could be viewed as the results one can obtain in the asymptotic case, samples $N_k \to \infty$, for the algorithms developed. We compare CDF based scheduling with PF scheduling [5] for some special versions of PF scheduling. For non opportunistic scheduling such as round robin, it is shown in [1] that CDF based scheduling achieves higher throughput for any user. An interesting question is whether CDF based scheduling can achieve the throughput of Opportunistic scheduling.

CDF based scheduling vs. Opportunistic scheduling

Proposition 1. *In a multiuser system with K users, CDF based scheduling can obtain the highest system throughput achieved by Opportunistic scheduling*[1]*, given the two scheduling techniques support the same allocation for the users, if and only if the distribution of users' throughput satisfy the following condition*

$$F_{R_k}(x) = F_{R_j}(x)^{\eta_{j,k}}, \tag{3.3}$$

for all $j, k = 1, \ldots, K$, where R_k, and R_j are the throughput of user k, j, defined by $R_k = \log_2(1 + \rho Z_k)$, with ρ being the transmit SNR, and $\eta_{j,k} = \frac{w_k}{w_j}$.

Proof. The detailed proof is provided in Appendix 3.7.1 □

Remark: Opportunistic scheduling selects user based on the maximum SNR and so achieves the highest system throughput. This proposition states that under the limited above mentioned condition, CDF based scheduling can match with Opportunistic scheduling in achieving the highest system throughput.

CDF based scheduling vs. Proportional Fair (PF)

To further motivate CDF scheduling methods, we compare them with PF scheduling. In PF scheduling, the BS select a user k based on

$$k^* = \arg\max_k \frac{R_k}{\bar{R}_k^\beta}, \tag{3.4}$$

where R_k is the instant throughput, $\bar{R}_k$ is the moving average throughput [5, 44, 38] of user k at time N_k, and β is an adjustable parameter. The time variable N_k also denotes the number of collected CQI. The average throughput is updated

$$\bar{R}_k = \begin{cases} \left(1 - \frac{1}{N_k}\right) \bar{R}_k[N_k - 1] + \frac{1}{N_k} R_k & k^* = k \\ \left(1 - \frac{1}{N_k}\right) \bar{R}_k[N_k - 1] & k^* \neq k \end{cases}, \tag{3.5}$$

[1] In this work, Opportunistic scheduling specifically means the users are selected to maximize the system throughput.

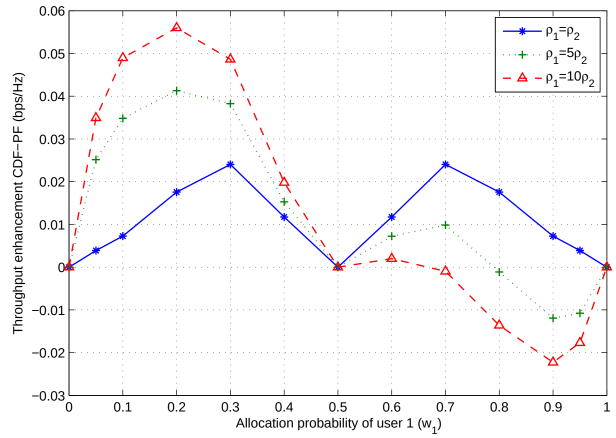


Figure 3.1: Relative performance of the CDF based scheduling in comparison to the modified PF scheduling under Rayleigh fading.

where $\bar{R}_k[N_k - 1]$ is the average throughput from the previous instant.

In order to analytically compare CDF based scheduling and PF scheduling, a modified version of PF scheduling as in [38] is used. We select the user with

$$k^* = \arg\max_k \frac{X_k}{\bar{X}_k^{\beta_k}}, \tag{3.6}$$

where X_k is the instantaneous SNR, $\bar{X}_k$ is the average SNR of user k instead of the average obtained over a certain time window.

To enable a fair comparison, we analyzed a simplified system with 2 users under Rayleigh fading. The allocation for the users are kept the same for both CDF based and PF scheduling[2]. Then, the analytical throughput of a system using CDF based scheduling and PF scheduling are compared in Fig. 3.1. In this figure, the difference in system throughput between CDF and PF scheduling as a function of resource allocation w_1 for user 1 are shown. The allocation for user 2 is $w_2 = 1 - w_1$. A positive value on the curve shows CDF scheduling is better and vice versa.

From the figure, it can be seen that CDF based scheduling outperforms PF scheduling for a large range of the weight w_1. When the average SNR of the

[2]The allocation for PF scheduling is achieved by properly selecting β.

two users are equal, CDF based scheduling is always better than or equal to PF scheduling. This simple experiment together with the previous analysis shows that CDF based scheduling is a viable alternative to PF scheduling and can be often superior when fairness is considered. We thus pursue a deeper understanding of CDF based scheduling and do expect similar results for more general heterogeneous systems. In summary, CDF based scheduling is found to have favorable properties and is worthy of further study.

3.3 Practical CDF based scheduling

Though CDF based scheduling is promising, it is based on the assumption that the CDF of CQI of every user is known at the transmitter. In practice, having knowledge of the CDF of the channel, even an approximate one, is challenging. Moreover, the channel model varies arbitrarily depending on user location, environment, and mobility. For these reasons, CDF based scheduling has been considered impractical.

These challenges motivate us to develop practical CDF based scheduling methods, which can be readily applied to real systems. Based on assumptions of the channel, we develop two classes of practical scheduling algorithms; Non-Parametric CDF based Scheduling (NPCS) and Parametric CDF based Scheduling (PCS). The algorithm NPCS is used when the fading model is not known at the BS. Meanwhile, the algorithm PCS is used when the BS knows the parameterized fading model and thus can simply learn the parameters of the channel.

In these algorithms, the BS collects CQI data, both the current and latest instantaneous CQI and the CQI from previous channel uses for each user, and makes use of the collected data. In this work, we assume the number of CQI samples collected for each of the users is different. This is mathematically more general and can be practically supported in an environment where user channels have different coherence times as well as activity levels. Thus, the BS obtains and keeps a different number of CQI samples N_k for each user $k = 1, \ldots, K$.

3.3.1 Non-Parametric CDF based Scheduling: Unknown Fading model

In Non-Parametric CDF based Scheduling (NPCS), the collected samples of channel CQI are used directly for the scheduling decision without trying to directly approximate the CDF of the channel.

We assume to possess N_k CQI samples from the current and previous channel uses for each user k and use these samples for the CDF based scheduling algorithm. A suitable value N_k might be chosen based on the channel dynamics of user k. In this work, we do not consider in detail how many samples have been selected but simply assume N_k before hand. These N_k samples are sorted in an ascending order $z_{k(1)} \leq \cdots \leq z_{k(N_k)}$. We also assume that the current CQI, denoted by z_k^I,, is in the i_k-th position in the ordered set, i.e. $z_k^I = z_{k(i_k)}$. If the CDF was known, then the random CQI variable Z_k would be mapped to a random variable U_k, uniformly distributed in $[0, 1]$, using the CDF, i.e. $U_k = F_{Z_k}(Z_k)$. Since the CDF is a nondecreasing function, the CDF mapping would result in the ordered CQI to map to an ordered set of values in the interval $[0, 1]$, i.e. the CDF mapping would result in $u_{k(i)} = F_{Z_k}(z_{k(i)})$ which are also ordered; $u_{k(1)} \leq \cdots \leq u_{k(i_k)} \leq \cdots \leq u_{k(N_k)}$. Since the order is preserved, the current CQI would map to $u_{k(i_k)}$, the i_k-th position and we use the indicator function $1_{u_{k(i_k)}}$ to denote that the most recent CQI has position i_k in the order statistics. Even though we do not know the exact values of each $u_{k(i_k)}$, we know their statistics [53], and can make use of them to develop a practical scheduler. To illustrate, one option is to use the average value $u_{k(i_k)}$ for the CDF scheduler leading to the following scheme.

ECS: Using the Expected value of the ordered CDF for Scheduling

- **Initialization:** For each user k, collect N_k CQI including the current instantaneous CQI z_k^I and the past $(N_k - 1)$ ones.

- Sort the CQI in an ascending order. Identify the position of the current CQI z_k^I, say i_k.

- Calculate the expected value[3] of the variable obtained by the CDF mapping of the ordered variable. This is given by $q_k = E\{U_{k(i_k)}\} = \frac{i_k}{N_k+1}$, $k = 1, .., K$. [53].
- Select[4] a user $k^* = \arg\max_k q_k^{\frac{1}{w_k}}$.
- **End.**

Using the expected value of the ordered CDF is simple and can be an option for CDF based scheduling. When we investigate the portion of resource allocated to the users, this method does not guarantee the desired resource allocation as shown in Fig. 3.4. We note that the CDF approximated to be $\frac{i-1}{N_k-1}$ in [51] is a bias estimation of the CDF. The mismatch in allocation for users does not reveal when every users have the same priority and collect the same number of samples. Because the mismatch between the desired and the actual allocation when ECS or algorithms in [51, 52] are used can be significant, we are more interested in algorithms which can provide better control over resource allocation.

For this purpose, we consider the construction of a mapping from $z_{k(i)}$ to the i-th element $\tilde{u}_{k(i)}$ preserving some statistical properties of $u_{k(i)}$, and ensure that the resultant random variable $\tilde{U}_k$ is a uniform random variable on the interval $[0, 1]$, mimicking the statistics of U_k. A minimum requirement on $f_{\tilde{u}_{k(i)}}(x)$, which denotes the density function of $\tilde{u}_{k(i)}$, is that it satisfy the following two conditions

$$\begin{aligned} &\int_0^1 f_{\tilde{U}_{k(i)}}(x)dx = 1, && \forall i = 1, \ldots, N_k,\ \forall k \quad (a) \\ &\frac{1}{N_k}\sum_{i=1}^{N_k} f_{\tilde{U}_{k(i)}}(x) = 1, && \forall x,\ \forall k \quad (b) \end{aligned} \tag{3.7}$$

where (a) comes from the properties of a density function and (b) comes from the fact that N_k equally likely ordered random variables $\tilde{U}_{k(i)}$ constitute an uniform random variable $\tilde{U}_k$ on the interval$[0, 1]$. Any mapping that satisfies this properties

[3] Instead of the mean, the median or any specifically manipulated value e.g. $E\left\{U_{k(i_k)}^{\frac{1}{w_k}}\right\}$ can be used.

[4] If two users have the same q_k, a tie breaking rule has to be implemented.

results in a mapping from the CQI to a uniform random variable on the interval $[0, 1]$ and can potentially be the basis of the NPCS algorithm as described below.

Class of Non-Parametric CDF based scheduling (NPCS) algorithms:

- For each of the K users generate a sample value to be used for resource allocation
  - **Initialization:** Collect N_k CQI includes the instantaneous CQI and the past ones.
  - Sort the CQI in an ascending order and identify the position of the instantaneous CQI, say i_k.
  - Generate the corresponding sample value $\tilde{u}_k$ for user k to be used for resource allocation.
    - Generate $\tilde{u}_{k(i_k)}$ a sample using the pdf $f_{\tilde{U}_{k(i_k)}}(\cdot)$ and set $\tilde{u}_k = \tilde{u}_{k(i_k)}$.
- Select a user $k^* = \arg\max_k \tilde{u}_k^{\frac{1}{w_k}}$.
  - Since the scheduling is being carried out using i.i.d uniform random variable, by properties of the CDF scheduler in [21], user k is ensured the proper fraction w_k of network resources.
- **End**

The above procedure meets the allocation criteria, but does not ensure maximization of multi-user diversity and hence system throughput. So the choice of the mapping $f_{\tilde{U}_{k(i_k)}}(x)$, must be chosen to optimize a desired objective function subject to the constraints in (3.7). We expand on this further.

Given a generated value $\tilde{u}_k = \tilde{u}_{k(i_k)}$, user k is selected if $\tilde{u}_j^{\frac{1}{w_j}} < \tilde{u}_{k(i_k)}^{\frac{1}{w_k}}$, $\forall j \neq k$. With a value of $\tilde{u}_{k(i_k)}$, probability user k is selected is $Pr\{k^* = k|\tilde{u}_{k(i_k)}; 1_{u_{k(i_k)}}\} = \prod_{j \neq k} \tilde{u}_{k(i_k)}^{\frac{w_j}{w_k}} = \tilde{u}_{k(i_k)}^{\frac{1}{w_k}-1}$. The probability user k is selected, given its channel's instantaneous CQI has position i_k-th, is calculated by taking expectation over value of

$\tilde{U}_{k(i_k)}$.

$$Pr\{k^* = k | 1_{u_{k(i_k)}}\} = \int_0^1 x^{\frac{1}{w_k}-1} f_{\tilde{U}_{k(i_k)}}(x)dx. \tag{3.8}$$

With some manipulations, the achievable throughput of the system can be shown to be

$$\begin{aligned} R_{NPCS} &= \sum_{k=1}^{K}\sum_{i=1}^{N_k} Pr\{1_{u_{k(i_k)}}\} Pr\{k^* = k|1_{u_{k(i_k)}}\} \int_0^\infty \log_2(1+x) f_{X_{k(i_k)}}(x)dx \\ &= \sum_{k=1}^{K}\sum_{i=1}^{N_k} \frac{1}{N_k} Pr\{k^* = k|1_{u_{k(i_k)}}\} R_{k(i_k)}, \end{aligned} \tag{3.9}$$

where $Pr\{1_{u_{k(i_k)}}\} = \frac{1}{N_k}$. The rate $R_{k(i_k)} = \int_0^\infty \log_2(1+x) f_{X_{k(i_k)}}(x)dx$ depends only on the distribution of $X_{k(i_k)}$ and does not depend on the mapping technique.

Considering the total system throughput as the objective function, one can optimize $f_{\tilde{U}_{k(i_k)}}(x)$, subject to (3.7). We now suggest some algorithms that are practical and show that they have desirable properties.

3.3.2 Some specific NPCS Algorithms

We now present two mappings that are based on the NPCS scheduler framework.

NPCS-1

From position i_k-th of the recent CQI on the ordered CQI sequence, the algorithm generates $\tilde{U}_{k(i_k)}$ with the same distribution as $U_{k(i_k)}$. Note that the density function of $U_{k(i_k)}$ is known because it is the i_k-th order statistic of N_k i.i.d random variables uniform on the interval $[0, 1]$. This mapping satisfies the constraints in

(3.7). From (3.8), the probability that user k is selected is

$$\begin{aligned} Pr\{k^* = k | 1_{u_{k(i_k)}}\} &= \int_0^1 x^{\frac{1}{w_k}-1} f_{\tilde{U}_{k(i_k)}}(x)dx \\ &= \int_0^1 N_k \binom{N_k-1}{i_k-1} x^{i_k+\frac{1}{w_k}-2}(1-x)^{N_k-i_k}dx \\ &= N_k \binom{N_k-1}{i_k-1} B\left(i_k + \frac{1}{w_k} - 1, N_k - i_k + 1\right), \end{aligned} \tag{3.10}$$

where the last equation follows from the definition of Beta function [54, 8.380].

Combined with (3.9), the average system throughput for NPCS-1 is

$$\begin{aligned} R_{NPCS-1} = &\sum_{k=1}^{K} \sum_{i_k=1}^{N_k} \binom{N_k-1}{i_k-1} B\left(i_k + \frac{1}{w_k} - 1, N_k - i_k + 1\right) \\ &\times \int_0^\infty \log_2(1+x) f_{X_{k(i_k)}}(x)dx. \end{aligned} \tag{3.11}$$

Our investigation into $NPCS-1$ shows that it does not achieve the highest system throughput. So we do not pursue its properties any further and use the algorithm to merely illustrate it as an example of a potential algorithm made possible by the NPCS framework.

NPCS-2

In this method, the random variable $\tilde{U}_k$ is created from $\tilde{u}_{k(i_k)}$ which is uniformly distributed in $[Q_{k,i_k-1}, Q_{k,i_k}]$. The boundary points Q_{k,i_k} are set equally spaced as in an uniform quantizer [55] as depicted in Fig. 3.2

$$Q_{k,i_k} = i_k \Delta_k, \tag{3.12}$$

for $i_k = 0, \ldots, N_k$ and $\Delta_k = \frac{1}{N_k}$. The values for $\tilde{u}_{k(i_k)}$ is generated employing a random variable uniformly distributed in $[Q_{k,i_k-1}, Q_{k,i_k}]$. It is easy to verify that this mapping satisfies the constraints in (3.7).

In NPCS-2, the following desired condition is guaranteed: If $i < j$ then the values generated will satisfy $\tilde{u}_{k(i)} < \tilde{u}_{k(j)}$. The ordering is preserved as would be

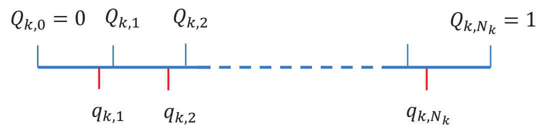


Figure 3.2: Mapping the i-th order CQI to the value $q_{k,i}$ on the empirical CDF.

the case if the true CDF was known. This is in contrast with NPCS-1.

Proposition 2. *The NPCS-2 mapping has the following properties*

1. *The expectation of $\tilde{U}_{k(i_k)}$, the i_k-th variable in the constructed ordered sequence has mean value*

$$E\{\tilde{U}_{k(i_k)}\} = \frac{2i_k - 1}{2N_k}, \tag{3.13}$$

2. *The NPCS-2 mapping is biased with the bias given by*

$$E[U_{k(i_k)} - \tilde{U}_{k(i_k)}] = \frac{N_k + 1 - 2i_k}{2N_k(N_k+1)} \xrightarrow[N_k \to \infty]{} 0 \tag{3.14}$$

3. *The variance of the difference between $U_{k(i_k)}$ and $\tilde{U}_{k(i_k)}$ is*

$$\begin{aligned} \sigma_{i_k}^2 &= E(U_{k(i_k)} - \tilde{U}_{k(i_k)})^2 \\ &= \frac{2 + 5N_k + N_k^2(4 + 12i_k - 12i_k^2) + N_k^3(1 + 12i_k)}{24N_k^2 + 60N_k^3 + 48N_k^4 + 12N_k^5} \xrightarrow[N_k \to \infty]{} 0. \end{aligned} \tag{3.15}$$

Proof. The detailed proof is provided in Appendix 3.7.2. □

Now we calculate the throughput of NPCS-2 based scheduler. From (3.8), given the auxiliary variable $\tilde{u}_{k(i_k)}$ is generated using a random variable uniformly distributed in $[Q_{k,i_k-1}, Q_{k,i_k}]$, the probability that user k is selected in NPCS-2 is

$$Pr\{k^* = k | 1_{u_{k(i_k)}}\} = \frac{1}{Q_{k,i_k} - Q_{k,i_k-1}} \int_0^1 x^{\frac{1}{w_k}-1} dx = N_k w_k \left[Q_{k,i_k}^{\frac{1}{w_k}} - Q_{k,i_k-1}^{\frac{1}{w_k}} \right]. \tag{3.16}$$

Using order statistic [53], the SNR of user k has the following distribution

$$f_{X_{k(i_k)}}(x) = N_k \binom{N_k-1}{i_k-1} F_{X_k}(x)^{i_k-1}[1 - F_{X_k}(x)]^{N_k-i_k} f_{X_k}(x), \tag{3.17}$$

where $F_{X_k}(x) = F_{Z_k}\left(\frac{x}{\rho_k}\right)$ as the SNR $x_k = \rho z_k$.

Plugging (3.8) and (3.17) into (3.9), the average system throughput is

$$R_{NPCS-2} = \sum_{k=1}^{K} w_k \sum_{i_k=1}^{N_k} \left[Q_{k,i_k}^{\frac{1}{w_k}} - Q_{k,i_k-1}^{\frac{1}{w_k}} \right] \int_0^\infty \log_2(1+x) f_{X_{k(i_k)}}(x)dx \tag{3.18}$$

From the equation, we observe that the throughput of each user k does not depend on other users but only depends on the understanding of user k of its own channel.

Theorem 2. *Among NPCS algorithms, NPCS-2 achieves the highest system throughput.*

Proof. In this problem, the objective function is system throughput, the variables are the mapping functions, and the constraints are as listed in (3.7).

$$\begin{aligned} \max_{U_{k(i)}} \quad & \sum_{k=1}^{K} \sum_{i=1}^{N_k} \frac{1}{N_k} R_{k(i)} \int_0^1 x^{\frac{1}{w_k}-1} f_{\tilde{U}_{k(i)}}(x)dx \\ s.t. \quad & \int_0^1 f_{\tilde{U}_{k(i)}}(x)dx = 1, \qquad \forall i = 1, \ldots, N_k, \ \forall k \\ & \frac{1}{N_k} \sum_{i=1}^{N_k} f_{\tilde{U}_{k(i)}}(x) = 1, \qquad \forall x, \ \forall k, \end{aligned} \tag{3.19}$$

where $R_{k(i)} = \int_0^\infty \log_2(1+x) f_{X_{k(i)}}(x)dx$. Note that $R_{k(i)}$ do not depend on the variables being chosen and is a function of the actual CDF of the CQI. By virtue of the properties of the order statistics and the monotonic mapping from the CQI to rate, we have $R_{k(1)} \leq R_{k(2)} \leq \cdots \leq R_{k(N_k)}$. Inspection of the optimization problem readily reveals that the optimization is independent for each user. So we just consider the problem for a given user k. From the allocation of user k, $\frac{1}{N_k} \sum_{i=1}^{N_k} \int_0^1 x^{\frac{1}{w_k}-1} f_{\tilde{U}_{k(i)}}(x)dx = \sum_{i=1}^{N_k} w_{k(i)} = w_k$, with $w_{k(i)} =$

$\frac{1}{N_k}\int_0^1 x^{\frac{1}{w_k}-1} f_{\tilde{U}_{k(i)}}(x)dx$. The resulting optimization problem is

$$\begin{aligned} \max_{w_{k(i)};i=1,\ldots,N_k} \quad & \sum_{i=1}^{N_k} R_{k(i)} w_{k(i)} \\ s.t. \quad & \sum_{i=1}^{N_k} w_{k(i)} = w_k. \end{aligned} \tag{3.20}$$

For any pair of the mapping orders i and j, with $j > i$, that satisfy the constraints in (3.19), we fix $w_{k(i)} + w_{k(i)} \triangleq w_{k(i,j)} = \text{const}$. Then, the objective function is $R_{k(i)}w_{k(i)} + R_{k(j)}w_{k(j)} = R_{k(i)}w_{k(i,j)} + \left(R_{k(j)} - R_{k(i)}\right) w_{k(j)}$ which is a monotonically increasing function w.r.t. $w_{k(j)}$ because $R_{k(j)} > R_{k(i)}$. With the constraint $w_{k(i,j)} =$ const, we find the mapping that maximizes $w_{k(j)}$. Let us choose a point $\hat{u} \in (0,1)$ such that $\int_0^{\hat{u}} \left(f_{\tilde{U}_{k(i)}}(x) + f_{\tilde{U}_{k(j)}}(x)\right) dx = 1$. This is always possible from the unit area property of the density functions involved. Now for any function $f_{\tilde{U}_{k(j)}}(x)dx$, we have the following relation

$$\begin{aligned} \int_0^1 f_{\tilde{U}_{k(j)}}(x) x^{\frac{1}{w_k}-1} dx &= \int_0^{\hat{u}} f_{\tilde{U}_{k(j)}}(x) x^{\frac{1}{w_k}-1} dx + \int_{\hat{u}}^1 f_{\tilde{U}_{k(j)}}(x) x^{\frac{1}{w_k}-1} dx && (a) \\ &\le \hat{u}^{\frac{1}{w_k}-1} \int_0^{\hat{u}} f_{\tilde{U}_{k(j)}}(x) dx + \int_{\hat{u}}^1 f_{\tilde{U}_{k(j)}}(x) x^{\frac{1}{w_k}-1} dx && (b) \\ &= \hat{u}^{\frac{1}{w_k}-1} \int_{\hat{u}}^1 f_{\tilde{U}_{k(i)}}(x) dx + \int_{\hat{u}}^1 f_{\tilde{U}_{k(j)}}(x) x^{\frac{1}{w_k}-1} dx && (c) \\ &\le \int_{\hat{u}}^1 \left(f_{\tilde{U}_{k(i)}}(x) + f_{\tilde{U}_{k(j)}}(x)\right) x^{\frac{1}{w_k}-1} dx, && (d) \end{aligned} \tag{3.21}$$

where (b) follows from $x \le \hat{u}$ in the first integral. The equation (c) follows from $\int_0^{\hat{u}} f_{\tilde{U}_{k(i)}}(x)dx + \int_0^{\hat{u}} f_{\tilde{U}_{k(j)}}(x)dx = 1$ and $\int_0^{\hat{u}} f_{\tilde{U}_{k(i)}}(x)dx + \int_{\hat{u}}^1 f_{\tilde{U}_{k(i)}}(x)dx = 1$. The last inequality comes from $\hat{u} \le x$ in the range $[\hat{u}, 1]$.

Thus, for any mapping function $f_{\tilde{U}_{k(j)}}(x)$, the objective function is improved by replacing it by the mapping $\left(f_{\tilde{U}_{k(i)}}(x) + f_{\tilde{U}_{k(j)}}(x)\right)$ in the range $[\hat{u}, 1]$. The equality holds when $\int_0^{\hat{u}} f_{\tilde{U}_{k(j)}}(x)dx = 0$ or there is a hard boundary between $\tilde{U}_{k(i)}$ and $\tilde{U}_{k(j)}$, i.e. the support of $f_{\tilde{U}_{k(i)}}(x)$ and $f_{\tilde{U}_{k(j)}}(x)$ do not overlap.

This rule applies for any pair and so it results in the support of $f_{\tilde{U}_{k(i)}}(x)$ and $f_{\tilde{U}_{k(j)}}(x)$ having no overlap for all $i \neq j$. This requirement coupled with the constraints in (3.7), restated in (3.19), result in algorithm NPCS-2. Hence, NPCS-2 is the optimal algorithm.

□

Corollary 1. *NPCS-2 is the optimal algorithm for any objective function which is monotonically increasing function of CQI.*

Proof. The proof is similar to that of Theorem 2. □

An example is the problem of minimizing the system bit error rate (BER). For a fixed rate, the BER is a monotonically decreasing function w.r.t the CQI order. This minimization problem is equivalent to maximizing negative of the BER. Following the same proof, NPCS-2 is the optimal algorithm for minimizing BER.

3.3.3 Performance of NPCS algorithms in Rayleigh fading

The average system throughput shown in (3.18), can be applied to any fading channel model. To provide an illustration of system performance, we consider all the links to the users to be under Rayleigh fading. Then, the distribution of the SNR can be represented

$$F_{X_k}(x) = 1 - e^{-x/\rho_k}, x > 0, \tag{3.22}$$

with $\rho_k = \rho c_k$ is the received SNR of the user k with $c_k = E\{Z_k\}$ is the pathloss from the BS to user k. Then, the distribution of the SNR in (3.17) is

$$f_{X_{k(i_k)}}(x) = N_k \binom{N_k-1}{i_k-1} \frac{1}{\rho_k} \sum_{l=0}^{i_k-1} \binom{i_k-1}{l} (-1)^l e^{-(N_k+l-i_k+1)x/\rho_k}, \tag{3.23}$$

which is substituted into (3.11) to obtain the average rate R_{NPCS-1} or into (3.18) to obtain the average rate R_{NPCS-2}. Using the integration in [54, 4.337.2] to get $\int_0^\infty e^{-\frac{l+1}{\rho_k}x} \log(1+x)dx = -\frac{\rho_k}{l+1} e^{\frac{l+1}{\rho_k}} Ei\left(-\frac{l+1}{\rho_k}\right)$, we have the following expressions for the average system throughput as summarized in the theorem.

Theorem 3. *In a multiuser system under Rayleigh fading with K users using NPCS algorithms, the overall system throughput is*

$$R_{NPCS-1} = \frac{1}{\ln 2} \sum_{k=1}^{K} \sum_{i_k=1}^{N_k} N_k \binom{N_k-1}{i_k-1} B\left(i_k + \frac{1}{w_k} - 1, N_k - i_k + 1\right) \times \sum_{l=0}^{i_k-1} \binom{i_k-1}{l} (-1)^{l+1} \frac{1}{N_k + l - i_k + 1} e^{\frac{N_k+l-i_k+1}{\rho_k}} Ei\left(-\frac{N_k + l - i_k + 1}{\rho_k}\right), \quad (3.24)$$

$$R_{NPCS-2} = \frac{1}{\ln 2} \sum_{k=1}^{K} w_k \sum_{i_k=1}^{N_k} \left[Q_{k,i_k}^{\frac{1}{w_k}} - Q_{k,i_k-1}^{\frac{1}{w_k}}\right] N_k \binom{N_k-1}{i_k-1} \times \sum_{l=0}^{i_k-1} \binom{i_k-1}{l} (-1)^{l+1} \frac{1}{N_k + l - i_k + 1} e^{\frac{N_k+l-i_k+1}{\rho_k}} Ei\left(-\frac{N_k + l - i_k + 1}{\rho_k}\right), \quad (3.25)$$

where $B(\cdot,\cdot)$ is beta function, $Ei(.)$ is the exponential integral function [54], and $Q_{k,i_k} = \frac{i_k}{N_k}$; w_k and ρ_k are correspondingly the assigned weight and the received SNR of user k.

3.4 Parametric CDF based Scheduling (PCS)

We assume a known parameterized form for the channel model of each user and the collected data is used to estimate the model parameters. The knowledge assumed is related to the type of fading and also might include a prior on the parameters, e.g. the distribution of the pathloss at the region. A maximum a posteriori probability (MAP) estimate [56] of the parameter $\boldsymbol{\rho}_k$ can be obtained by maximizing the posterior density $f_{\boldsymbol{\rho}_k|\boldsymbol{X}_k}(\boldsymbol{\varrho}_k|\boldsymbol{x}_k)$. By Bayes rule

$$f_{\boldsymbol{\rho}_k|\boldsymbol{X}_k}(\boldsymbol{\varrho}_k|\boldsymbol{x}_k) = \frac{f_{\boldsymbol{X}_k|\boldsymbol{\rho}_k}(\boldsymbol{x}_k|\boldsymbol{\varrho}_k) f_{\boldsymbol{\rho}_k}(\boldsymbol{\varrho}_k)}{f_{\boldsymbol{X}_k}(\boldsymbol{x}_k)}. \quad (3.26)$$

and the estimate $\hat{\boldsymbol{\varrho}}_k$ is obtained by maximizing the posteriori probability density function

$$\left.\frac{\partial L(\boldsymbol{\varrho}_k, \boldsymbol{x}_k)}{\partial \boldsymbol{\varrho}_k}\right|_{\boldsymbol{\varrho}_k=\hat{\boldsymbol{\varrho}}_k} = \frac{\partial \log f_{\boldsymbol{X}_k|\boldsymbol{\rho}_k}(\boldsymbol{x}_k|\boldsymbol{\varrho}_k)}{\partial \boldsymbol{\varrho}_k} + \left.\frac{\partial \log f_{\boldsymbol{\rho}_k}(\boldsymbol{\varrho}_k)}{\partial \boldsymbol{\varrho}_k}\right|_{\boldsymbol{\varrho}_k=\hat{\boldsymbol{\varrho}}_k} = \mathbf{0} \qquad (3.27)$$

Algorithm PCS

- For each user, estimate the model parameters
  - **Initialization:** Collect N_k CQI including the current instantaneous CQI and the past values for user k.
  - Find a MAP estimate of the model parameters $\hat{\boldsymbol{\rho}}_{\boldsymbol{k}}$ of the channel model for user k.
- Select a user $k^* = \arg\max_k F_{X_k|\hat{\boldsymbol{\rho}}_k}(x_k|\hat{\boldsymbol{\varrho}}_{\boldsymbol{k}})^{\frac{1}{w_k}}$.
- **End.**

The proposed PCS algorithm is quite general. The parameter $\boldsymbol{\rho}_k$ is in general a vector of parameters needed to define the channel model. For illustration purposes, we now consider a Rayleigh fading channel.

PCS in Rayleigh fading

In a Rayleigh fading channel, the parameter ρ_k is just a scalar. $f_{X|\rho_k}(x|\varrho_k) = \frac{1}{\varrho_k}e^{-\frac{x}{\varrho_k}}u(x)$. Assuming i.i.d measurements and noninformative prior, the MAP estimate satisfies

$$\left.\frac{\partial L(\varrho, x)}{\partial \varrho}\right|_{\varrho=\hat{\varrho}_k} = -\frac{N_k}{\hat{\varrho}_k} + \sum_{n=1}^{N_k} \frac{x_n}{\hat{\varrho}_k^2} = 0, \qquad (3.28)$$

which results in the estimation formula as stated in Proposition 3.

Proposition 3. *Under Rayleigh fading, the estimated pathloss of a user k with* *weight w_k using PCS is*

$$\hat{\rho}_k = \frac{\sum_{n=1}^{N_k} x_n}{N_k}. \quad (3.29)$$

where $x_n, n = 1, .., N_k$, are the N_k collected data. The average throughput of the user is

$$R_{PCS} = \frac{1}{\ln 2} \int_0^\infty \sum_{k=1}^{K} \sum_{l=0}^{\infty} \binom{\frac{1}{w_k}-1}{l} (-1)^{l+1} \frac{N_k^{N_k}}{\Gamma(N_k)} e^{-\frac{N_k \hat{\rho}_k}{\rho_k}} \times \frac{\hat{\rho}_k^{N_k}}{\rho_k^{N_k} (\hat{\rho}_k + l\rho_k)} e^{\frac{l\rho_k + \hat{\rho}_k}{\rho_k \hat{\rho}_k}} Ei\left[-\frac{l\rho_k + \hat{\rho}_k}{\rho_k \hat{\rho}_k}\right] d\hat{\rho}_k, \quad (3.30)$$

where ρ_k is the average received SNR of user k, and $\binom{\frac{1}{w_k}-1}{l-1} = \frac{\prod_{m=1}^{l}\left(\frac{1}{w_k}-m\right)}{(l-1)!}$.

Proof. The detailed derivation for the average throughput is relegated to Appendix 3.7.3. We observe that in Rayleigh fading, the channel parameter is the pathloss which is estimated by averaging over N_k CQI samples as in (3.29). □

Impact of imperfect estimation

The resource allocation as well as the average throughput of a user k depends on the precision of the estimated pathloss ρ_k. We now investigate the impact of this estimation error.

Theorem 4. *Under Rayleigh fading, the probability that the difference between allocation probability of a user k and its weight w_k is smaller than ϵ is given by*

$$Pr\{|Pr\{k^* = k\} - w_k| < \epsilon\} = \frac{1}{\Gamma(N_k)} \times \begin{cases} \gamma\left(N_k, \frac{N_k}{1-\frac{\epsilon}{A_k}}\right) - \gamma\left(N_k, \frac{N_k}{1+\frac{\epsilon}{A_k}}\right) & A_k > \epsilon \\ 1 - \gamma\left(N_k, \frac{N_k}{1+\frac{\epsilon}{A_k}}\right) & A_k \leq \epsilon \end{cases}, \quad (3.31)$$

with $A_k = \left|-w_k + \sum_{l=0}^{\infty} \binom{1/w_k - 1}{l} (-1)^l \frac{1}{(l+1)^2}\right|$ and $\gamma(.)$ is an incomplete gamma function.

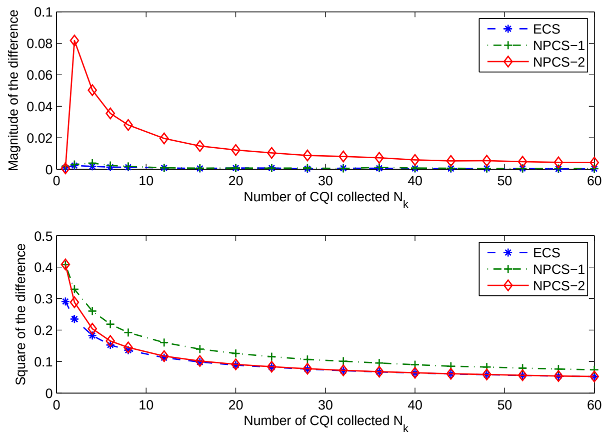


Figure 3.3: The precision of the mapping from ordered CQI to the quantized and interpolated CDF - the top figure is the average of the magnitude difference and the lower figure is the variance of the difference between the real CDF value with that of the proposed mapping methods.

Furthermore,

$$Pr\left\{|Pr\{k^* = k\} - w_k| < \epsilon\right\} \xrightarrow[N_k \to \infty]{} 1.$$

Proof. The detailed proof is in Appendix 3.7.4. □

Herein, the theorem shows the PCS algorithm can ensure an arbitrarily good precision in the resource allocation for users given we collect a reasonably large number of channel samples.

3.5 Simulation results

To evaluate the analytical results as well as the performance of the proposed approaches, we consider a multiuser system with $K = 10$ users where user k is assigned weights $w_k = (k+1)a_k$, where a_k is the normalization parameter to guarantee that $\sum_{k=1}^{K} w_k = 1$. The link from the BS to the users are under Rayleigh fading with the pathloss of each user k determined by $c_k = be^{-\lambda k}$ with $\lambda = 0.1$ and b is a constant so $\sum_{k=1}^{K} c_k = K$. The transmit SNR at the BS is set $\rho = 10$dB.

In Fig. 3.3, the precision of the mappings developed is evaluated. The

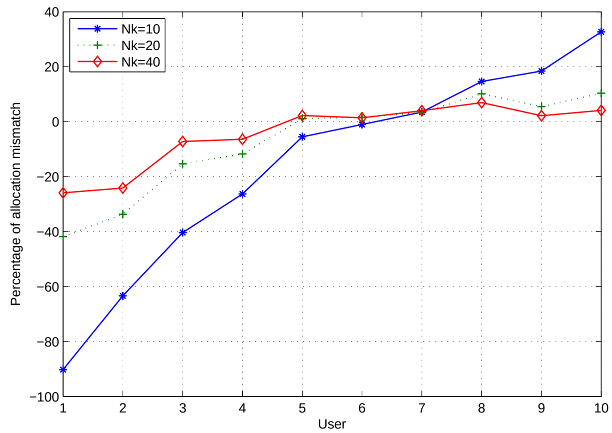


Figure 3.4: The mismatch, $\frac{Pr\{k^*=k\}-w_k}{w_k}$ (%), in resource allocation when the expected CDF is used.

metric used for the evaluation are $|E(\tilde{U}-U)|$ and $\text{var}(\tilde{U}-U)$, where $\tilde{U}$ and U are the estimated and actual random variables. The first metric is a measure of unbiasedness and the second is a measure of the mean squared error. Both ECS and $NPCS-1$ are unbiased estimates and so have smaller $|E(\tilde{U}-U)|$. When the means squared error (MSE) is used as a metric, NPCS-2 has smaller MSE than NPCS-1. However, both have higher MSE than ECS. This is because of the randomization introduced in these methods in order to generate a uniform random variable whereas ECS generates a discrete random variable. Both the average error and the MSE decrease as the number of samples N_k increase and go to zero when $N_k \to \infty$.

We note that ECS results in the smallest average error and MSE. However, this method does not guarantee the desired resource allocation for the users. Simulation results documenting the mismatch are shown in Fig. 3.4. When $N_k \to \infty$ we expect the mismatch to diminish. However, for finite N_k, the mismatch is significant and is unpredictable being heavily dependent on system parameters. This shows that if resource allocation has to be tightly controlled, the interpolation step is needed in NPCS algorithms. Also, the system throughput with the ECS scheduler is slightly better than NPCS-1 but poorer than NPCS-2 in our experiments. We also noticed the system throughput using ECS in comparison to the NPCS

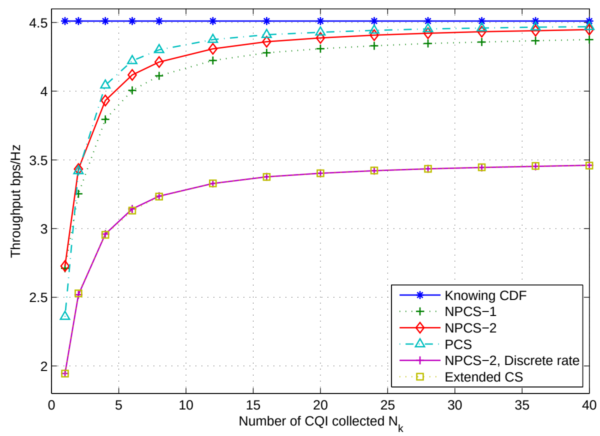


Figure 3.5: Performance of practical CDF scheduling as a function of collected channel data N_k under Rayleigh fading.

algorithms does not have a consistent pattern. The relative system throughput depends on system parameters such as number of users, allocation weights, and pathloss of these users etc making it difficult to draw conclusions. However, we are interested in methods with a strict requirement on the resource allocation for users and so only pay attention to algorithms NPCS-1 and NPCS-2.

The performance of the proposed scheduling methods as a function of number of samples N_k is shown in Fig. 3.5. From the figure, it can be seen that NPCS-2 obtains higher system throughput than NPCS-1 does. Furthermore, with information about the channel model, PCS utilize this fact and performs slightly better than the NPCS algorithms. When N_k increases, the achieved system throughput increases as expected since we obtain more accurate information about the channels of the users. With N_k large enough, e.g. $N_k > 30$ samples, the loss in throughput is under 0.7% in comparison to the case when the channel distribution is known perfectly. This bodes well for CDF based methods as this requirement is typically easy to obtain in real networks where the fast fading is of the order of milliseconds and the channel model does not change significantly in seconds. For example, if a CQI sample is collected in every 1-millisecond LTE frame, a thousand samples will be collected in 1 second which easily enables NPCS-2 to approach the performance of knowing perfectly the CDF of the channel.

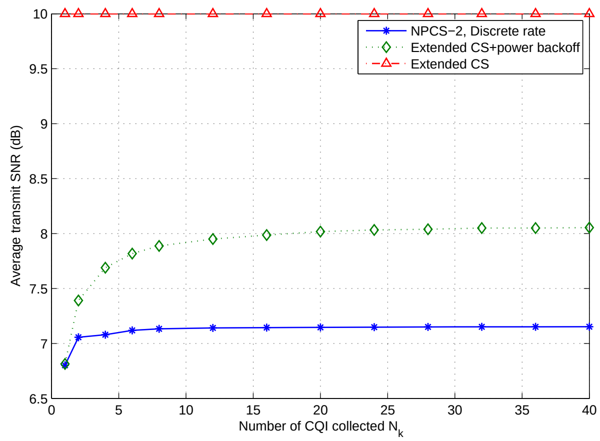


Figure 3.6: Average transmit SNR used in NPCS-2 and extended CS [1] as a function of collected channel data N_k under Rayleigh fading.

We now compare NPCS-2 with a CDF based technique proposed in [1] to support discrete rates. To enable the comparison, we use NPCS-2 to support a set discrete rates with $\{1, 2, 4, 6, 8\}$ bps/Hz which might correspond to BPSK, QPSK, 16QAM and so on. The two algorithms achieve the same system throughput as shown in Fig. 3.5. Because in NPCS-2 we use order statistics and employ a finer partition[5] of the resource allocation decision variable $\tilde{U}$, given a chosen rate, the algorithm has better selection and favors the user with a better channel quality. Given any chosen rate, if the system employs back off for the transmit SNR, i.e. power control, then NPCS-2 requires less average SNR for the transmission as shown in Fig. 3.6. The transmit power requirement is less than the extended CS in [1] by about 1 dB in our experiment setup assuming the extended CS has explicit CQI and is able to back off the transmit power. Equivalently, if the two algorithms have the same average SNR, NPCS-2 achieves about 10% higher system throughput. A more thorough comparison and study of the methods developed in this paper to support discrete rates is the subject of our future work. The preliminary results are very encouraging.

Typically, the fading model at the location of a user k depends on the geographical characteristic of the transmission environment and can be learnt and

[5]The partition in [1] is based on the number of discrete rates and in NPCS-2 is based on N_k.

characterized by the network. PCS exploits the learnt model and estimate the instantaneous parameters of the model. From Fig. 3.5, it is shown PCS has better performance than NPCS algorithms. Though the advantage of PCS might need further study, it is reasonable that having more understanding of the channel model can help improve allocation decisions.

The performance of CDF and PF scheduling are compared in Fig. 3.7. To ease the comparison, we consider a system with two users with the difference in the average received SNR being 10dB. In one set of simulations, both users are assume to be experiencing Rayleigh fading. In the other, one user is assumed under Rayleigh fading while the other is under Nakagami-m fading [57], with $m = 4$. The performance using PF scheduling with the parameter β in (3.4) is investigated first. For a chosen β, the allocation probability for the users, which is collected by averaging over 10^6 experiments, is used to set the corresponding weights in CDF based scheduling, e.g. weight of user 1 is set to equal $Pr\{k^* = 1\}$ in PF scheduling. We note that though the CDF based scheduling with perfect channel knowledge does not depend on N_k, its performance is not a constant in Fig. 3.7 because weights for users which is taken based on the performance of PF scheduling method, change slightly. When β is smaller than ~ 0.6, the PF scheduling is close to Opportunistic scheduling which favors and almost always (with probability $> 90\%$ in this experiment) allocates resource to the user with higher SNR. In this case, PF scheduling is better than CDF scheduling. When priorities of users are stressed more, it is observed that CDF scheduling outperforms PF scheduling. This situation happens frequently when we want to allocate comparable amounts of resource to the users. Similar results are observed with other combinations of fading types.

As the allocation of a user does not depend on parameters of others as well as number of user K, we consider a user k with weight $w_k = 0.5$. The resource allocation as a function of N_k for this user is investigated in the proposed learning method. Here, our concern is with the difference between the allocation probability for user k and the desired allocation probability w_k. The goal is to keep this discrepancy small, e.g. within range $[w_k - \epsilon, w_k + \epsilon]$. We note that the average allocation probability is reasonably guaranteed with a sufficient number of samples N_k. Specifically, with $N_k > 8$ the largest different between w_k and $Pr\{k^* = k\}$ is smaller than 0.005 in the experiments presented in Fig. 3.5. As N_k increases, the allocation probability for user k converges to w_k, as $Pr\{|Pr\{k^* = k\} - w_k| < \epsilon\}$ gradually goes to 1 as proved in theorem 4 and observed in Fig. 3.8.

3.6 Conclusion

We have proposed practical approaches to enable the application of CDF based scheduling technique in heterogeneous multiuser systems. The proposed NPCS and PCS algorithms are shown to precisely control resource allocation for users, simple enough to be employed in real systems, and frequently have better performance than the existing PF scheduling. In the comparison with PF scheduling, the CDF based scheduling is better when the fairness among users is of major concern. For the methods developed, the achievable throughput of each user increases and quickly approaches the throughput achievable with perfect knowledge of the CDF. Specifically, the throughput is about 99% of the perfect CDF case when the user has about 30 i.i.d channel samples making the methods practically viable.

The text of this chapter, in full, is a reprint of the paper [58], Anh H. Nguyen, Yichao Huang and Bhaskar D. Rao Learning methods used for CDF scheduling in multiuser heterogeneous systems, that has been submitted for publication on IEEE Transaction on Signal Processing. The dissertation author is the primary researcher and author, and the co-authors listed in this publication collaborated and supervised the research constitutes this chapter.

3.7 Appendices

3.7.1 CDF based scheduling versus Opportunistic Scheduling

We denote the PDF and CDF of the throughput from the BS to the users by $f_{R_k}(.)$ and $F_{R_k}(.)$ respectively with $k = 1, 2$. In Opportunistic scheduling, a user k is selected when it obtains the higher rate, the CDF of the rate of the selected users has the distribution $F_{R_{k^*}}(x) = F_{R_1}(x)F_{R_2}(x)$ [53]. The average system throughput using Opportunistic scheduling R_O is

$$R_O = \int_0^\infty xd[F_{R_1}(x)F_{R_2}(x)]. \tag{3.32}$$

In CDF scheduling, the CDF of the rate of the selected user k^* is $F_{R_{k^*}}(x)^{\frac{1}{w_k}}$. Referred from [49], the average system throughput using CDF based scheduling R_{CDF} is

$$R_{CDF} = \int_0^\infty xd[w_1 F_{R_1}(x)^{\frac{1}{w_1}} + w_2 F_{R_2}(x)^{\frac{1}{w_2}}]. \tag{3.33}$$

Then, we want to prove $R_{CDF} \leq R_O$ and work out the condition for the equality.

$$R_{CDF} - R_O = \int_0^\infty xd[w_1 F_{R_1}(x)^{\frac{1}{w_1}} + w_2 F_{R_2}(x)^{\frac{1}{w_2}} - F_{R_1}(x)F_{R_2}(x)] \tag{3.34}$$

We consider the function $F_{R_i}(x)^{\frac{1}{w_i}} = e^{y_i}$ which is a convex function w.r.t $y_i = \frac{1}{w_i}\ln F_{R_i}(x)$. From the properties of convex functions, we have $\alpha e^{y_1} + (1-\alpha)e^{y_2} \geq e^{(\alpha y_1 + (1-\alpha)y_2}$. Set $\alpha = w_1$, we have

$$w_1 e^{y_1} + w_2 e^{y_2} \geq e^{(w_1 y_1 + w_2 y_2)},$$

or equivalently, by replacing $y_i = \frac{1}{w_i}\ln F_{R_i}(x)$, we obtain

$$g(x) = w_1 F_{R_1}(x)^{\frac{1}{w_1}} + w_2 F_{R_2}(x)^{\frac{1}{w_2}} - F_{R_1}(x)F_{R_2}(x) \geq 0. \tag{3.35}$$

We now prove $\int_0^\infty xdg(x) < 0$. Note that $\lim_{x\to 0} g(x) = 0$ and $\lim_{x\to\infty} g(x) = 0$. We have $\int_0^\infty xdg(x) = xg(x)|_0^\infty - \int_0^\infty g(x)dx$. Let us begin with the achievable rate of a user k, $\int_0^\infty x f_{R_k}(x)dx$ which is finite. Then, $\lim_{x\to\infty} F_{R_k}(x) > 1 - x^{-\alpha}$, for some $\alpha > 1$. Otherwise, we have $\int_0^\infty f_{R_k}(x)xdx > Pr\{R_i > x\}x = \lim_{x\to\infty}(1 - F_{R_k}(x))x = \infty$ which contradicts with the finite rate condition. As a consequence, $x\left(F_{R_k}(x)^{\frac{1}{w_k}} - F_{R_k}(x)\right) = O(x^{1-\alpha})$ which goes to zero when x approach infinity. The term $w_k x\left[F_{R_k}(x)^{\frac{1}{w_k}} - F_{R_k}(x)\right]\Big|_0^\infty = 0$. Hence, $\lim_{x\to 0} xg(x) = 0$ and $\lim_{x\to\infty} xg(x) = 0$ or equivalently $xg(x)|_0^\infty = 0$ which result in

$$\int_0^\infty xdg(x) = -\int_0^\infty g(x)dx < 0, \tag{3.36}$$

because $g(x) > 0$. Hence, CDF based scheduling can not outperform Opportunistic scheduling, which is not a surprise result. Our interest is to determine under which conditions, the two scheduling methods have equal performance. From (3.35), the two scheduling methods have similar performance when the equality holds, which results in $g(x) = 0$. This happens when $y_1 = y_2$ or

$$F_{R_1}(x)^{\frac{1}{w_1}} = F_{R_2}(x)^{\frac{1}{w_2}} \ \forall x. \tag{3.37}$$

Equivalently, CDF scheduling has the same selection as Opportunistic scheduling or the user with the higher SNR is selected. In a system with K users, CDF scheduling has the same selecting result as Opportunistic scheduling for any couple of users. In the end, the user with the highest SNR is selected which means the algorithm obtains highest system throughput.

3.7.2 The precision of the mapping in NPCS algorithm

The expectation of the i_k-th variable in the constructed ordered sequence of CDF as described in NPCS algorithm, is

$$\begin{aligned} E\{U_{k(i_k)}\} &= \int_0^1 N_k \binom{N_k-1}{i_k-1} x^{i_k-1}[1-x]^{N_k-i_k} x dx & (a) \\ &= N_k \binom{N_k-1}{i_k-1} B(i_k+1, N_k-i_k+1) & (b) \\ &= \frac{i_k}{N_k+1}, & (c) \end{aligned} \tag{3.38}$$

where we get (a) from the PDF of the variable as in (3.17) and by utilizing the fact the CDF of a variable uniformly distributed in $[0,1]$ is $F_U(x) = x$. Equation (b) comes from [54, 8.380] and the last equation is due to the definition of Beta function [54, 8.384.1].

Similar, the variance of $U_{k(i_k)}$ can be calculated

$$\begin{aligned} \sigma^2_{U_{k(i_k)}} &= E\{U^2_{k(i_k)}\} - E\{U_{k(i_k)}\}^2 \\ &= \int_0^1 N_k \binom{N_k-1}{i_k-1} x^{i_k-1}[1-x]^{N_k-1} x^2 dx - \frac{i_k^2}{(N_k+1)^2} \\ &= \frac{i_k(N_k+1-i_k)}{(N_k+1)^2(N_k+2)}, \end{aligned} \tag{3.39}$$

where $E\{U^2_{k(i_k)}\} = \frac{i_k(i_k+1)}{(N_k+1)(N_k+2)}$ which is similar to the calculation in deriving the mean.

NPCS-2

The mapping from the unknown CDF of the CQI to a variable $\tilde{U}_{k(i_k)}$ is bias. The average difference is

$$E[U_{k(i_k)} - \tilde{U}_{k(i_k)}] = \frac{i_k}{N_k+1} - \frac{1}{2}\left[\frac{i_k}{N_k} + \frac{i_k-1}{N_k}\right] = \frac{N_k+1-2i_k}{2N_k(N_k+1)} \xrightarrow[N_k\to\infty]{} 0 \tag{3.40}$$

The variance of the difference between $U_{k(i_k)}$ and $\tilde{U}_{k(i_k)}$ is obtained

$$\begin{aligned}\sigma^2_{U_{k(i_k)}-\tilde{U}_{k(i_k)}} &= E\{(U_{k(i_k)}-\tilde{U}_{k(i_k)})^2\} - E\{(U_{k(i_k)}-\tilde{U}_{k(i_k)})\}^2 \\ &= E\{U^2_{k(i_k)}\} - 2E\{U_{k(i_k)}\}E\{\tilde{U}_{k(i_k)}\} + E\{\tilde{U}^2_{k(i_k)}\} - E\{(U_{k(i_k)}-\tilde{U}_{k(i_k)})\}^2 \\ &= \frac{2+5N_k+4N_k^2+12i_kN_k^2-12i_k^2N_k^2+N_k^3+12i_kN_k^3}{24N_k^2+60N_k^3+48N_k^4+12N_k^5}, \end{aligned} \tag{3.41}$$

where $E\{\tilde{U}_{k(i_k)}\} = \frac{1}{2}\left[\frac{i_k}{N_k} + \frac{i_k-1}{N_k}\right]$ and $E\{\tilde{U}^2_{k(i_k)}\} = \frac{i_k^3-(i_k-1)^3}{3N_k^2}$. When $N_k \to \infty$, we have

$$\lim_{N_k\to\infty} \sigma^2_{u_{k(i_k)}-\tilde{u}_{k(i_k)}} = \lim_{N_k\to\infty} \frac{i_k + \frac{1}{12} - \frac{i_k^2}{N_k}}{N_k^2} = 0. \tag{3.42}$$

3.7.3 Performance of PCS

From (3.45), the probability that a user k with channel SNR x is selected is $F_{X_k(\hat{\rho}_k)}(x)^{\frac{1}{w_k}-1}$. The average throughput of the user is

$$\begin{aligned} R_{PCS|\hat{\rho}_k} &= \int_0^\infty \log_2(1+x) F_{X_k(\hat{\rho}_k)}(x) f_{X_k}(x)dx \\ &= \int_0^\infty \frac{1}{\rho_k}\sum_{l=0}^{\infty}\binom{\frac{1}{w_k}}{l}(-1)^l e^{-x\frac{l\rho_k+\hat{\rho}_k}{\rho_k\hat{\rho}_k}}\log_2(1+x)dx \\ &= \frac{1}{\ln 2}\sum_{l=0}^{\infty}\binom{\frac{1}{w_k}-1}{l}(-1)^{l+1}\frac{\hat{\rho}_k}{\hat{\rho}_k+l\rho_k} e^{\frac{l\rho_k+\hat{\rho}_k}{\rho_k\hat{\rho}_k}} Ei\left[-\frac{l\rho_k+\hat{\rho}_k}{\rho_k\hat{\rho}_k}\right], \end{aligned} \tag{3.43}$$

where $f_{X_k}(x) = \frac{1}{\rho_k}e^{-\frac{x}{\rho_k}}$ and $F_{X_k(\hat{\rho}_k)}(x) = 1 - e^{-\frac{x}{\hat{\rho}_k}}$. The last equation comes from the exponential integral defined in [54, 4.337.2].

From the estimated pathloss $\hat{\rho}_k = \frac{1}{N_k}x_{k,i}$ where $x_{k,i}$ is the collected CQIs,

$\hat{\rho}_k$ has $\chi^2_{2N_k}$ distribution. The average system throughput is

$$\begin{aligned}
R_{PCS} &= \frac{1}{\ln 2}\int_0^\infty R_{PCS2|\hat{\rho}_k} f_{\hat{\rho}_k}(\hat{\rho}_k) d\hat{\rho}_k \\
&= \frac{1}{\ln 2}\int_0^\infty \sum_{k=1}^{K}\sum_{l=0}^{\infty} \binom{\frac{1}{w_k}-1}{l} (-1)^{l+1} \frac{N_k^{N_k}}{\Gamma(N_k)} e^{-\frac{N_k \hat{\rho}_k}{\rho_k}} \\
&\quad \times \frac{\hat{\rho}_k^{N_k}}{\rho_k^{N_k}(\hat{\rho}_k + l\rho_k)} e^{\frac{l\rho_k+\hat{\rho}_k}{\rho_k\hat{\rho}_k}} Ei\left[-\frac{l\rho_k+\hat{\rho}_k}{\rho_k\hat{\rho}_k}\right] d\hat{\rho}_k. \qquad (3.44)
\end{aligned}$$

where $f_{\hat{\rho}_k}(\hat{\rho}_k) = \frac{N_k}{\rho_k \Gamma(N_k)} \left(\frac{N_k \hat{\rho}_k}{\rho_k}\right)^{N_k-1} e^{-\frac{N_k \hat{\rho}_k}{\rho_k}}$. As we do not have a close form expression for the integral, we use numerical integration for calculating the average throughput of the system using PCS algorithm. To avoid possible error amplification due to the infinite sum in (3.44), we prefer to apply the numerical integration after taking the sum.

3.7.4 The precision in resource allocation of PCS

As the main concern is a user k, perfect CDF knowledge is assumed for other users. Then, user k, with the instantaneous SNR x is selected if the weighed version of its calculated CDF is greater than weighted CDF of any other user j

$$F_{X_j(\rho_j)}(x_j)^{\frac{1}{w_j}} \leq F_{X_k(\hat{\rho}_k)}(x)^{\frac{1}{w_k}}, \qquad (3.45)$$

where x_j is the SNR on channel of user j. The probability user k is selected is

$$\begin{aligned}
Pr\{k^* = k\} &= \int_0^\infty \prod_{j\neq k} F_{X_k(\hat{\rho}_k)}(x)^{\frac{w_j}{w_k}} f_{X_k}(x) dx \\
&= \int_0^\infty F_{X_k(\hat{\rho}_k)}(x)^{\alpha_k} f_{X_k}(x) dx, \qquad (3.46)
\end{aligned}$$

where $\alpha_k = \frac{1}{w_k} - 1$. Note that $F_{X_k(\hat{\rho}_k)}(x) = 1 - e^{-\frac{x}{\hat{\rho}_k}}$ is the calculated CDF based on the estimated parameter $\hat{\rho}_k$ and $f_{X_k}(x) = \frac{1}{\rho_k} e^{-\frac{x}{\rho_k}}$ is the real distribution of the

SNR of user k. Using Taylor expansion we have the following approximation

$$\begin{aligned} f_{X_k}(x) &= f_{X_k(\hat{\rho}_k)}(x) + f'_{X_k(\hat{\rho}_k)}(x)\Delta\rho_k + O\left((\Delta\rho_k)^2\right) \\ &\simeq f_{X_k(\hat{\rho}_k)}(x)\left[1 + x\frac{\Delta\rho_k}{\hat{\rho}_k^2} - \frac{\Delta\rho_k}{\hat{\rho}_k}\right], \end{aligned} \tag{3.47}$$

The precision of the approximation is shown in Fig. 3.9. When the difference of ρ_k and $\hat{\rho}_k$ is not large, the approximation is almost equal to the real function. where $\Delta\rho_k = \rho_k - \hat{\rho}_k$. Then, the probability that user k is selected is

$$\begin{aligned} Pr\{k^* = k\} &= \left(1 - \frac{\Delta\rho_k}{\hat{\rho}_k}\right)\int_0^\infty F_{X_k(\hat{\rho}_k)}(x)^{\alpha_k} f_{X_k(\hat{\rho}_k)}(x)dx \\ &+ \frac{\Delta\rho_k}{\hat{\rho}_k^2}\int_0^\infty F_{X_k(\hat{\rho}_k)}(x)^{\alpha_k} f_{X_k(\hat{\rho}_k)}(x)xdx \\ &= w_k\left(1 - \frac{\Delta\rho_k}{\hat{\rho}_k}\right) + \frac{\Delta\rho_k}{\hat{\rho}_k}\int_0^\infty [1-e^{-y}]^{\alpha_k} e^{-y} y dy \\ &= w_k - w_k\frac{\Delta\rho_k}{\hat{\rho}_k} + \frac{\Delta\rho_k}{\hat{\rho}_k}\int_0^\infty \sum_{l=0}^{\infty} \binom{\alpha_k}{l} e^{-(l+1)y}(-1)^l y dy \\ &= w_k + \frac{\Delta\rho_k}{\hat{\rho}_k}\left(-w_k + \sum_{l=0}^{\infty}\binom{\alpha_k}{l}(-1)^l\frac{1}{(l+1)^2}\right), \end{aligned} \tag{3.48}$$

where $y = \frac{x}{\hat{\rho}_k}$ and we use Taylor expansion to obtain the third equation. To quantify how many samples are needed so the allocation for user k is precise enough, we consider the distribution of $\hat{\rho}_k$. As $\hat{\rho}_k = \sum_{l=1}^{N_k} \frac{x_l}{N_k}$ where each random variable x_l has chi-squared distribution with 2 degrees of freedom, $N_k\frac{\hat{\rho}_k}{\rho_k}$ has chi-squared distribution with $2N_k$ degrees of freedom.

$$F_{\hat{\rho}_k}(\hat{\rho}_k) = \frac{1}{\Gamma(N_k)}\gamma\left(N_k, \frac{\hat{\rho}_k N_k}{\rho_k}\right) \tag{3.49}$$

Then, if we want the resource allocation for user k has the precision ϵ, which means $|Pr\{k^* = k\} - w_k| < \epsilon$ or equivalently

$$\hat{\rho}_k^{(l)} < \hat{\rho}_k < \hat{\rho}_k^{(r)}, \tag{3.50}$$

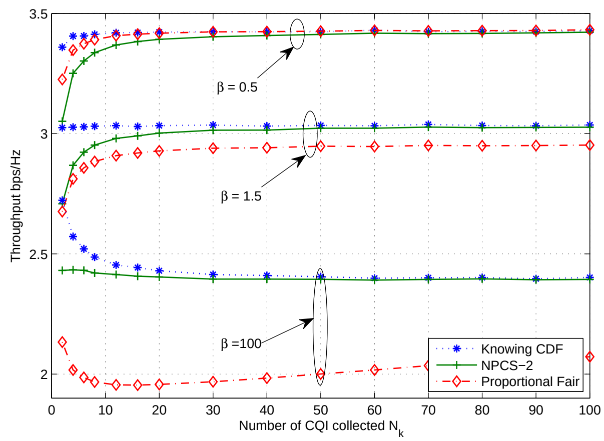
with $A_k = \left|-w_k + \sum_{l=0}^{\infty} \binom{\alpha_k}{l} (-1)^l \frac{1}{(l+1)^2}\right|$. The $\hat{\rho}_k^{(l)} = \frac{\rho_k}{1+\frac{\epsilon}{A_k}}$ while $\hat{\rho}_k^{(r)} = \frac{\rho_k}{1-\frac{\epsilon}{A_k}}$ if $A_k > \epsilon$ and $\hat{\rho}_k^{(r)} = \infty$ if $A_k < \epsilon$. Then, this condition $Pr\left\{|Pr\{k^* = k\} - w_k| < \epsilon\right\}$ needs to be guaranteed to happen with high probability, we have the result in (3.31).

From this equation, we consider a chi-square distribution random variable with the same CDF denoted ν, with $2N_k$ degree of freedom. Then, ν has mean N_k and variance $2N_k$. From (3.31), set $\Delta_w = \min\left(N_k - \frac{N_k}{1+\frac{\epsilon}{A_k}}, \frac{N_k}{1-\frac{\epsilon}{A_k}} - N_k\right) = \frac{N_k \epsilon}{A_k+\epsilon}$, we have $Pr\left\{|Pr\{k^* = k\} - w_k| < \epsilon\right\} \geq Pr\left\{N_k - \Delta_w < \nu < N_k + \Delta_w\right\}$. Using Chebyshev inequality, we have

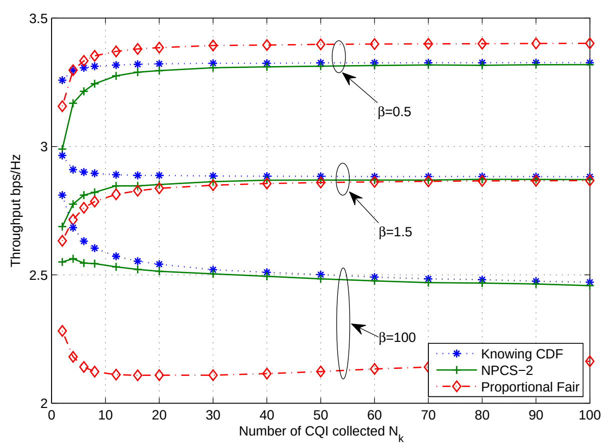
$$Pr\{|\nu - N_k| > m\sqrt{2N_k}\} \leq \frac{1}{m^2}. \tag{3.51}$$

Choosing $m = \frac{\Delta_w}{\sqrt{2N_k}} = \sqrt{N_k}\frac{\epsilon}{\sqrt{2}(A_k+\epsilon)}$, combined with (3.31), we obtain

$$Pr\left\{|Pr\{k^* = k\} - w_k| < \epsilon\right\} \geq 1 - \frac{2}{N_k}\left(1 + \frac{A_k}{\epsilon}\right)^2 \xrightarrow[N_k \to \infty]{} 1. \tag{3.52}$$



a) Rayleigh fading



b) Rayleigh and Nakagami-m fading

Figure 3.7: Comparison between CDF and PF scheduling as a function of collected channel data N_k. In the first model, (a), both users are under Rayleigh fading and in the second model, (b), one user is under Rayleigh fading and the other user is under Nakagami-m fading with $m = 4$.

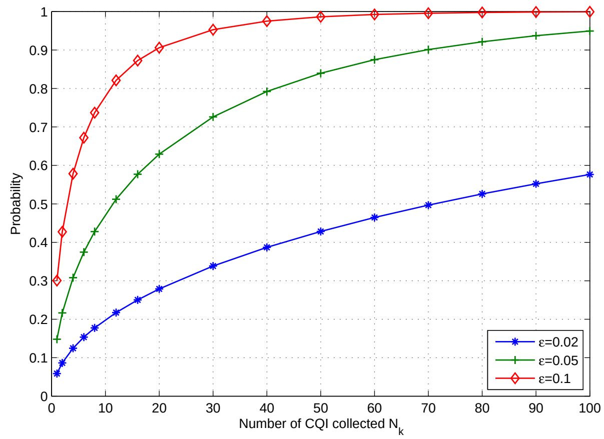


Figure 3.8: Allocation precision $Pr\{|Pr\{k^* = k\} - w_k| < \epsilon\}$ as a function of collected channel data N_k under Rayleigh fading.

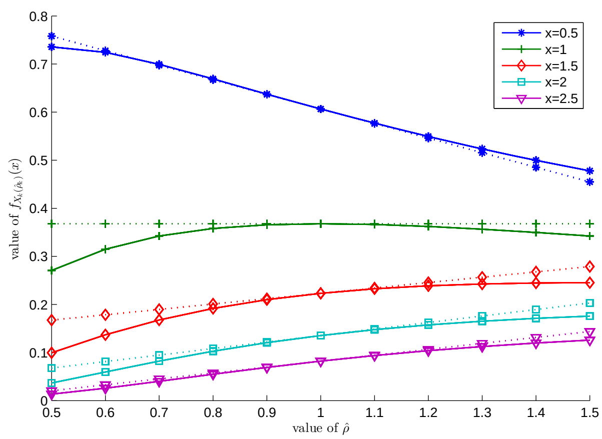


Figure 3.9: An illustration of the approximation of $f_{X_k(\rho_k)}(x)$ by $f_{X_k(\hat{\rho}_k)}(x)$ as in (3.47) where the continuous lines are real values and the discontinuous lines are for the approximated values.

Chapter 4

CDF scheduling algorithms for finite rate multiuser systems with limited feedback

In this work, simple and practical CDF scheduling methods are developed that preserve the virtues of CDF scheduling, namely fairness and effective use of multiuser diversity. They are the accelerated extended CDF scheduling (AeCS), modified non-parametric CDF scheduling (modified NPCS), and the CDF scheduling with optimized Quantizers (CSwQ) methods. The developed methods do not need a priori knowledge of the channel distribution and can work effectively in systems that support discrete number of transmission rates and are constrained by limited feedback resources. They exploit the limited feedback resource for their scheduling decisions and transmission rate selections and learn the distribution of channel quality information of each user, as needed, in order to best exploit multiuser diversity. The CSwQ method is shown to have desirable properties in terms of both system throughput and power savings. Analytical results for the system throughput and power savings of the CSwQ method are developed that are valid for a finite number of observation samples. Then, through numerical simulations, the developed methods are shown to effectively save transmit power and achieve high system throughput. In addition, they converge rapidly with the number of data samples to the ideal case wherein perfect knowledge of the channel CDF is assumed.

4.1 Introduction

A significant challenge in wireless communication system design is providing satisfactory service to a large number of users with varying needs and conditions. This can be partially interpreted as achieving both the highest system throughput and reasonably meeting each individual requirement. The first goal of high system throughput is achieved by being opportunistic and exploiting multiuser diversity [2]. The second goal of meeting user needs is achieved by sacrificing system throughput to maintain fairness among users [5]. Specifically, the system allocates resources to users with poorer channels to guarantee their needs are met even though such allocations are not supportive of achieving high system throughput. Because these goals are in conflict, good scheduling methods need to support fairness among users with minimal loss in overall system throughput. There are many scheduling methods and in this paper we consider cumulative distribution function (CDF) based scheduling proposed in [21]. This is because of the many promising attributes of the CDF scheduling, namely flexible and precise control of temporal fairness among users while at the same time ensuring high system throughput. A requirement and challenge in CDF scheduling is that the CDF of each user's channel quality has to be known a priori. The practical scheduling methods developed in [51, 58] do not require a priori knowledge of channel distribution but learn them from the data and have been shown to retain many of the advantages of CDF scheduling wherein the CDF is assumed to be known perfectly. Motivated by the attractive features of CDF scheduling, in this work we seek to further advance and modify the developed methods so that they can be applied effectively to real systems with limited feedback resources and that support a finite number of transmission rates.

As mentioned above, ensuring satisfactory quality of services for all users is of utmost importance in wireless networks. To do so, many fairness preserving scheduling methods have been developed which attempt to guarantee that every user is adequately served while efficiently utilizing network resource by exploiting multiuser diversity. Some commonly used fairness preserving schedulers are temporal fairness [4], game theory based fairness [59], utilitarian fairness [60, 8, 61], Max-Min fairness [62, 63] which maximizes the minimum among rates of the users, and max rate scheduling which are special cases of proportional fairness (PF) [5, 6, 7]. Fairness can also be supported by using utility functions [43], and by minimizing potential delay as in [26]. On one end of the fairness versus throughtput tradeoff spectrum is round robin scheduling [3], where users are served in a sequential manner and hence fairness is guaranteed. However, the approach does not exploit multiuser diversity and therefore results in poor overall system throughput. A popular and effective method is proportional fairness (PF) [5]. The method selects a user based on both the instantaneous channel quality as well as the average rate experienced in the past which is measured over a suitably chosen time window. However, analytical tractability of the PF scheduling in a multiuser environment with different user requirements is challenging and intractable to date.

This motivates our work on CDF scheduling [21, 49, 50]. Interestingly, CDF based scheduling has superior ability to precisely control temporal fairness among users and has the capability to make a good tradeoff between diversity and fairness. Recent studies have shown that the method is tractable in fairly complex environments and the analysis can lead to interesting system insights. The CDF-based scheduling is leveraged in a general multicell network in [48] and in a partial feedback OFDMA relay system in [49] to guarantee scheduling fairness and simultaneously obtain multiuser diversity gain. In [50], the CDF-based scheduling is analytically studied in the random beamforming framework, and the notion of individual sum rate and individual scaling laws are proposed to characterize the performance under this scheduling policy. The challenge of learning the CDF and making them practical is addressed in [1, 51, 58]. The ranking based scheduling method in [51, 52] works very well and is able to ensure all users in the system are served equally. However, the performance when users have different priorities and different number of CQI samples is unclear. The non-parametric CDF scheduling (NPCS) method[1] is shown to be optimal for maximizing overall system throughput. Also, through numerical simulations, the practical CDF scheduling methods developed are shown to be frequently superior to PF scheduling particularly when fairness is emphasized [58]. However, the methods developed in [51, 58] assume no limitation on feedback resources and also assume that the system can support a continuum of rates. In [1], a system that supports a finite number of rates is considered. An interesting method, the Extended Cumulative distribution function Scheduler (ECS), is developed and analyzed. However, the method developed converges very slowly, which limits it use, and also feedback resources are considered in a limited context.

In CDF based scheduling, similar to other opportunistic scheduling methods, the BS needs channel information to calculate the chosen scheduling metric for user selection and resource allocation. The channel information is hence vital to the performance of these scheduling methods and is obtained through a feed-

[1]Because NPCS-2 has the best performance among NPCS algorithms proposed in [58], we use NPCS to refer to NPCS-2.

back mechanism [64]. The type of channel information that the BS has depends on the feedback method employed. An overview of feedback methods, the importance of feedback information, the state of the art techniques in feedback reduction as well as many related issues is provided in [23]. With feedback information, multiuser diversity can be exploited [22, 24] and maximum system throughput can be achieved [2]. The overall throughput is shown in [25] to have double logarithmic growth not only for full channel state information (CSI) but also for one bit feedback systems. Though it is most desirable to have the highest quality feedback information [65], such perfect feedback information is not available in real systems. In fact, any type of imperfection, which can be due to measurement error or feedback delay [66, 67, 68, 69], results in degradation in system performance which can be reflected in lower overall throughput, higher packet error rate [66] or more frequent outage conditions. In reality, only quantized feedback information is available [70] and the feedback can be optimized to maximize the average rate. The impact of limited feedback is further studied in [71, 72, 73, 74], wherein the system performance is analyzed and transmission techniques are suitably adjusted. Based on practical constraints, in this work we consider a system with limited feedback. Specifically, quantized feedback is assumed where each user is allocated a fixed number of feedback bits.

The work developed in this paper addresses the challenge of CDF scheduling methods in the context of systems with limited number of transmission rates (M) and limited feedback resources per user (B). The outline and contributions of this work are as follows. In section II, the multiuser system model is presented along with the background on CDF scheduling. In section III, a modification to the extended CS [1], the accelerated extended CS (AeCS) method, is proposed to address the slow convergence problems of the extended CS method. To address the issues of feedback bits and varying number of training samples per user, the modified NPCS method [58] is proposed. Interestingly, this modified method together with a back-off scheme requires much less power than AeCS. As a consequence, the method achieves significantly higher throughput if the two use the same average transmit power. However, if the two have the same maximum transmit power, this method has a slight lower throughput in comparison with AeCS. We propose a CDF Scheduler with optimized Quantizers (CSwQ) which combines the virtues of both the modified NPCS and the AeCS. As a result, the proposed CSwQ scheme has the highest throughput while using significantly less power than the extended CS scheme. To best adapt the feedback strategy to each user's channel distribution, a method to fine tune the user's feedback quantizer according to the collected CQI samples is proposed. Finally, the throughput and power consumption of the CSwQ is analyzed. In contrast to the asymptotic nature of the results in [1], the performance analysis results are valid for a finite number of samples. Then, in section IV, experimental results are presented to support and verify the performance of the proposed methods. The simulations are shown to agree well with the results. Furthermore, the system throughput is shown to converge quite rapidly to the optimal throughput when the CDF is assumed known.

4.2 System model

We consider a multiuser downlink system with K users, each equipped with a single antenna. These users are served by a base station (BS) which also has a single antenna. Though the downlink system is specifically considered, a similar consideration can be applied to the uplink. Upon knowing the channel information, the BS selects a user for resource allocation. The received signal at a user k is

$$y_k = h_k \sqrt{\rho} s_k + n_k, \tag{4.1}$$

where y_k is the received signal, s_k is the transmit signal, and $n_k \sim CN(0, 1)$ is the additive noise. ρ is the transmit power and referred as transmit SNR since the noise is fixed at variance one. $h_k \in C^{1\times 1}$ is the channel from the BS to the selected user k, which is assumed to be independent over time, Throughout this work, the channels as well as noise are assumed i.i.d. across users and over time.

We denote the random variable associated with the CQI and the SNR of a user k by capital letters Z_k and X_k, where the instantaneous values are $z_k = |h_k|^2$ and $x_k = \rho z_k$. To select a user to be served on a resource block, CDF based scheduling [21] is used. The CDF scheduler selects a user as follows:

$$k^* = \arg\max_k u_k^{\frac{1}{w_k}}, \text{ where } u_k = F_{Z_k}(z_k). \tag{4.2}$$

w_k is the weight which represents the proportion of network resource allocated to user k. The positive weight w_k can be understood as priority of user k and is preassigned to all users $k = 1, \ldots, K$ so that $\sum_{k=1}^{K} w_k = 1$. The transformation of the CQI Z_k to $U_k = F_{Z_k}(Z_k)$ is an important step in computing the metric for scheduling. For discussion purposes, we use the terminology U-space when working with the transformed random variables and the CQI or original space when working with the CQI. Several advantages follow from the CDF scheduling policy. Note that the entire distribution plays a role in scheduling and that the random variable U_k obtained from Z_k is a uniform random variable between $[0, 1]$. All the CQI are transformed to uniform random variables and so they appear identically distributed to the scheduler. From the user perspective, they also have the perception of competing with other users with similar characteristics! The other user distributions do not impact the performance of a particular user, in contrast to the PF scheduler, greatly simplifying the resource allocation and service guarantee task. Secondly, because the CDF is used in the transformation, and the CDF is a monotonically increasing function[2], large values in CQI also results in large values in the transformed random variable. The selection policy ensures that the channel is assigned to the user k when the channel of user k is in a favorable condition, i.e. user with the minimum $P(Z_k > z_k)$ is scheduled. In other words, a user is scheduled when the user's channel supports a rate which is high enough but least probable to become higher, i.e. high end tail of its rate distribution. Thirdly, the resource needed can be allocated and readily met by choosing the weights, i.e. $P(k^* = k) = w_k$. Fourthly, this scheduler lends itself well for analytical studies and for developing insights as explained previously.

However, in real systems, the channel distribution needed to calculate $F_{Z_k}(z_k)$ in (4.2) is usually unknown. CDF based scheduling methods for finite

[2]This is for simplicity. The general non-decreasing property can also be readily accommodated with some caveats

rate systems with limited feedback resources and when the channel distribution is unknown are developed in this work.

4.3 Finite rate CS scheduling

In CDF scheduling, the BS upon selecting a user can transmit at the maximum rate supported by the link. In practice, the BS can only transmit at one of some finite predefined rates [1]. The rate options for the BS are assumed be a set of M rates $r_0 = 0, \ldots, r_{M-1}$. It is assumed that the BS selects rate r_i for communicating with a user if the capacity of the link to that user is greater than r_i but smaller than r_{i+1}.

Now we examine the relationship between the number of transmission rates M, the number of feedback bits per user B and the number of collected CQI per user N_k. The number of transmission rates M is typically small, and we assume that the number of feedback bits B assigned to each user to be more than capable of encoding these transmission rates. The BS collects N_k CQI samples from each user and with time, and reasonable channel coherence time, N_k can be large. This leads to the relationship

$$M \leq 2^B \ll N_k. \tag{4.3}$$

The problem we discuss is how to design an effective CDF based scheduler under these constraints.

4.3.1 Window based Extended CS

An interesting scheme, the extended CS, was proposed in [1]. Though feedback bits were not explicitly mentioned, it can be interpreted as a scheme where $M = 2^B$. We briefly summarize the scheme to motivate our work.

Algorithm 1. *Extended CS*

- *Each user k measures its CQI and reports the rate index i_k.*

- *The scheduler draws a sample* u_k *from a uniform random variable* U_k *in the interval* $[q_{k,i_k-1}, q_{k,i_k}]$ *and transforms it into a scheduling metric* $u_k^{\frac{1}{w_k}}$ *for user* k*.*

- *The scheduler selects a user whose scheduling metric is the largest, as in (4.2).*

- *The BS transmit to user* k^**.*

- *The scheduler updates the probabilities* $p_{k,i}$ *of the rates, and* $q_{k,i}$ *the corresponding partition in the* U *space as follows*

$$p_{k,i} \leftarrow \lambda p_{k,i} + (1-\lambda) 1_{i=i_k}$$
$$q_{k,i} \leftarrow \sum_{j=1}^{i} p_{k,j}, \tag{4.4}$$

where $0 < \lambda < 1$*.*

In this method, λ is a parameter that controls the convergence rate of the method. An interesting feature of the method is that asymptotically, on convergence, it partitions the interval $[0, 1]$ (U space) into M disjoint regions in an optimal manner. The feedback requirements are minimal and equal $\log M$ bits. However, there is room for improvement. The convergence rate of the method is quite slow as shown in the numerical experiments presented in [1]. As a result, the partitions can be quite poor in an environment with short coherence times. So more optimal use of the data would be beneficial in learning the partitions. The number of feedback bits usually can be larger than the rates and it would be desirable to extend the method to accommodate this possibility. Since the number of rates are fixed, the improvement, from the additional number of bits, is more likely to be in better user selection and power savings. We will elaborate more on this later.

Accelerated Extended CS (AeCS): For Extended CS, we found the convergence rate can be improved by making more aggressive use of the data initially and by setting the parameter λ adaptively, i.e. time varying λ_k. At the instant N_k, the parameter λ is set as follows: $\lambda_k = \frac{N_k-1}{N_k}$. This approach is akin to a decreasing step size in adaptive methods with the weight given to new samples diminishing with time, i.e. $1-\lambda_k = \frac{1}{N_k}$. This strategy is appropriate in a stationary environment but not for time varying environments. To deal with the non-stationary (slowly time varying) case, one can simply choose not to reduce λ_k below a predetermined threshold. We do not pursue this in this paper. Our simulations show that this simple change greatly speeds up the convergence rate of the methods making it viable in practical scenarios. However, this does not address how to more effectively exploit feedback resources when $2^B > M$. For the rest of the paper, we are interested in developing methods which can converge fast as well as effectively adapt with the availability of feedback resources.

4.3.2 Modified NPCS Algorithm

To better utilize feedback resources and to support finite rate transmission, we consider adaptation of a recently developed NPCS algorithm for learning the CDF [58]. In this method, the N_k samples are sorted in an ascending order $z_{k(1)} \leq \cdots \leq z_{k(N_k)}$. The order statistics based approach appears more natural because large values of CQI are likely to play a more important role in scheduling and density mapping at the high end is more relevant. If the CDF was known, then the random CQI variable Z_k would be mapped to a random variable U_k, uniformly distributed in $[0,1]$, using the CDF, i.e. $U_k = F_{Z_k}(Z_k)$. Since the CDF is a nondecreasing function, the CDF mapping would result in the ordered CQI to map to an ordered set of values in the interval $[0,1]$, i.e. the CDF mapping would result in $u_{k(i)} = F_{Z_k}(z_{k(i)})$ which are also ordered; $u_{k(1)} \leq \cdots \leq u_{k(i_k)} \leq \cdots \leq u_{k(N_k)}$. Since the order is preserved, the current CQI, assumed to be in order i_k would map to $u_{k(i_k)}$, the i_k-th position. Note that even though the order statistics of the CQI are not known, the distribution of the order statistics in the U space, $U_{k(i)}$, are well known. Since the mean of the ordered statistics of $U_{k(i)}$ are known, NPCS considers a construction of a mapping from $z_{k(i)}$ to the i-th element $\tilde{u}_{k(i)}$ by partition the interval $[0,1]$ into disjoint intervals with the mean values as the centers and generating a random variable uniform over that interval. As shown in [58], NPCS with explicit feedback is the optimal algorithm for continuous rate transmission. For limited feedback systems, it is still optimal if the feedback resource is adequate to convey the order of user's CQI, i.e. when $B > \log_2(N_k)$. However, in practice $B < \log_2(N_k)$ would be the regime of interest, and it is modified to accommodate this constraint. The algorithm is described as follows.

Algorithm 2. *Modified NPCS algorithm*

- ***CDF mapping:*** *For each of K users, generate a sample value to be used for resource allocation.*
  - *Initialization:*
    - *The CDF domain $[0, 1]$ is divided into $Q = 2^B$ equally spaced bins. The bin i is the interval $[q_{k,i}, q_{k,i+1}] = \left[(i-1)\frac{1}{Q}, i\frac{1}{Q}\right]$ with $1 \leq i \leq Q$.*
    - *Each user k collects N_k CQI samples which includes the instantaneous CQI and the past ones. These CQI are sorted in an ascending order and the position of the instantaneous CQI j is identified.*
  - *Identify a feedback index*
    - *Calculate the expected value of the CDF, given the order j [58]*

$$E\{U_{k,j}\} = \frac{j}{N_k + 1}, \tag{4.5}$$

      with $1 \leq j \leq N_k$.
    - *Identify the feedback index i_k if $\frac{j}{N_k+1} \in [q_{k,i_k}, q_{k,i_k+1})$.*
  - *The BS upon receiving the feedback index from a user k, generates the corresponding sample value $\tilde{u}_k$ based on a uniform distribution over the interval $[q_{k,i_k}, q_{k,i_k+1})$, where q_{k,i_k} is the boundary between bins i_k and $i_k + 1$.*
- ***User selection:*** $k^* = \arg\max_k \tilde{u}_k^{\frac{1}{w_k}}$.
- ***Rate selection:*** *The selected user feeds back its explicit CQI for rate selection.*

As can be seen from the description of the algorithm, the algorithm is quite simple and the computation and data collection tasks can be conducted in a distributed manner. Each user keeps its own channel data and generates a feedback index in each scheduling interval. The BS collects feedback indexes from all users, generates the corresponding CDF samples and makes a scheduling decision. An advantage of this approach is that the computational requirements for the BS as well as for each user are minimal. In contrast to Extended CS, which is rate centric, this algorithm is CQI centric and hence power centric. As a result, an advantage of this approach is that the selected user, though is not guaranteed to have the highest rate, has the higher average SNR since the user is more likely to be selected when it's channel is in good condition. This implies that the user is likely to have better CQI than needed for the discrete rate selected and supported by the system. Hence the BS can reduce transmit power to meet the SNR requirements for that rate, thereby leading to power savings. Since the fedback procedure does not incorporate rate information, in order for the BS to transmit outage free, the user must feed back the rate index upon selection. This step is an extra requirement. Furthermore, though the overall SNR is much better than that in AeCS, as shown by the numerical results in Section 4.4 (Fig. 4.2), there is a slight loss in overall throughput as observed in the experiments [75]. This can be attributed to the fact that the boundaries in the U space generated by modified NPCS based on CQI do not coincide with the rate boundaries and so some regions can encompass two rates. We therefore look for a method that can make the best of AeCS and modified NPCS, and can effectively address both the loss in throughput of the modified NPCS and the low overall SNR of AeCS.

4.3.3 CDF scheduling with optimized Quantizers (CSwQ)

In the proposed scheme, we suggest doing the partition of the U-space in a hierarchical manner. First partition the $[0, 1]$ interval into M disjoint intervals corresponding to the rates. This is an attempt to capture the virtues of AeCS. Then partition each rate region more finely accounting for the remaining $2^B - M$ quantization levels. This is an attempt to capture the virtues of NPCS. Though these finer regions are not beneficial in term of increasing system throughput, it helps in the selection of an user with favorable channel conditions and hence potential power savings. For this discussion we assume that the number of partitions for each rate region, L_i for rate r_i, is fixed and known a priori. Intuitively, rates more likely to be supported by the channel should be partitioned more finely and this makes use of the feedback resources more effectively. We will discuss the choice of L_i later. We denote $q_{k,i}$ as the empirical CDF corresponds to the quantization level i. The notation $q_{k,i,j}$ corresponds to the empirical CDF in the jth sub-interval, given the interval i is sub-divided into many sub-intervals. Furthermore, for this discussion, we assume all N_k samples have been collected. A data adaptive version can be developed but is not discussed. The proposed scheme is as follows:

Algorithm 3. *CSwQ*

- ***CDF mapping:*** *For each of K users, generate a sample value to be used for resource allocation.*
  - *Initialization:*
    - *Define feedback regions and M corresponding indexes.*
      - *Determine M rate regions $[0, \ldots, r_{M-1}, \infty]$. This is a one to one mapping from the SNR to the rates.*
      - *Divide each rate interval i into L_i intervals. For this discussion, we assume an uniform partition in the rate domain as follows: $r_{i,j} = r_i + j\frac{r_{i+1}-r_i}{L_i}, j = 1, .., L_i$. The total number of encoded regions is $Q = 2^B$ and hence*

$$\sum_{i=1}^{M} L_i = Q. \tag{4.6}$$

      - *The set L_i, with $i = 1, \ldots, M$ is assumed known[3].*
    - *Each user k collects N_k CQI samples including the instantaneous CQI and the past ones.*

[3]The optimized value of L_i is set and agreed beforehand between the BS and each user k as described in algorithm 4.

- *Identify a feedback index which jointly encodes the rate index i_k as well as the index of the finer partition j, i.e.$\{i_k, j\}$*
    - *Identify the rate index i_k of the instantaneous CQI.*
    - *Identify the sub-region j with $j = 1, \ldots, L_{i_k}$, that the instantaneous CQI belongs to*[4].
- *The BS generates a corresponding sample value $\tilde{u}_k$ for user k, using a random variable which is uniformly distributed in $[q_{k,i_k,j-1}, q_{k,i_k,j}]$ with $q_{k,i_k,j}$ is the empirical CDF corresponds to the level i_k, j. If the level i_k, j is defined by the rate $r_{i_k,j}$, then $q_{k,i_k,j}$ is the percentile of samples with CQI smaller than the CQI corresponded to level i_k, j. The interval $[q_{k,i_k,j-1}, q_{k,i_k,j}]$ is set as described in Alg. 4.*

- ***User selection:*** $k^* = \arg\max_k \tilde{u}_k^{\frac{1}{w_k}}$.
- ***Rate selection:*** *The selected user k transmits with rate r_{i_k}.*

In Alg. 3, the procedure to map the position of the current CQI in the CQI sequence to a feedback index $\{i_k, j\}$, and then how to use this index for scheduling decision, is described. The feedback quantizer for each user k is assumed to be known before hand. In a heterogeneous network, these quantizers should be customized in accordance to each user's channel characteristic. Herein, we describe a learning method to customize the quantizer of each user based on the collected data for that user.

Algorithm 4. *Constructing quantizers in the CSwQ*

- *Each user k collects N_k CQI samples which includes the instantaneous CQI and the past ones.*
- *From the set of transmission rates $r_0, \ldots, r_{i_k}, \ldots, r_{M-1}$, assume $m_{k,i}$ samples fall in the rate region between rate $[r_i, r_{i+1})$. As the system has more feedback resource to encode each CQI sample, each rate interval i can be divided into*

[4] Similarly, the rate region i_k can be divided equally in CDF domain as described in algorithm 4.

L_i sub-intervals [5]*. There are two options for dividing a rate region i into L_i smaller regions.*

- *Equal space in CDF (U) domain: $q_{k,i,j} = q_{k,i} + j\frac{q_{k,i+1}-q_{k,i}}{L_i}$ with $j = 1,\ldots,L_i - 1$* [6]*.*
- *Equal space in rate domain. $r_{i,j} = r_i + j\frac{r_{i+1}-r_i}{L_i}$ with $r_{i,j}$ denotes the rate in jth sub-interval in the interval between the transmission rate r_i and r_{i+1}.*

- *The value of L_i is determined so that $L_i \geq 1$ and each interval has roughly the same number of samples. Let's assume we divide rate region $[r_i, r_{i+1})$ into equally space sub-intervals in rate domain. If we find any rate intervals a and b such that $\frac{m_{k,a}}{L_a+1} \geq \frac{m_{k,b}}{L_b-1}$, we increase L_a and decrease L_b by 1: $L_a \leftarrow L_a + 1$ and $L_b \leftarrow L_b - 1$. Then, the boundaries $q_{k,i_k,j}$ in Alg. 3 is the percentile value based on the samples of user k that support a rate smaller than $r_{i_k,j}$.*

Similar to the application of the NPCS method, this scheme is also capable of distributing the computation and data collection tasks among the BS and all users. The difference is a quantizer has to be agreed upon between the BS and each user k. There are two ways to have such agreement. One way is the quantizer can be optimized by each user using Alg. 4 and fedback to the BS. The BS and each user k use that quantizer until a change is needed. The other is more distributed wherein the BS and each user run Algorithm 4 independently and obtain the same quantizer. At first, they can use a default quantizer which only sets rate boundaries. Upon counting the number of samples in each rate region, the BS and each user k can decide to further divide each rate region into finer sub-regions in a predetermined manner. As they use the same procedure, the number of sub-regions in each rate region are the same for both user k and the BS. This method results in more computational overhead at the BS because it has to compute quantizers for all users. However, the advantage is no information exchange on quantizers is

[5] The constructed quantizer is identical for each user.

[6] To divide the right most rate region, we assume r_M is selected so that the probability the link can have rate higher than r_M is negligible.

needed between the BS and users. The quantizers can be updated more frequently based on the arriving data samples.

4.3.4 Performance of CSwQ

To evaluate the proposed method, we investigate its performance given the number of feedback bits B and the number of transmission rate M, under the assumption that $B \geq \log_2 M$. When $B = \log_2 M$ this method is identical to AeCS. In [1], an analysis for extended CS is presented ($B = \log_2 M$) which shows system performance when the CDF boundary of $\tilde{u}_k$ is stable. For this condition to hold, the parameter λ has to be set arbitrarily close to 1, which in turn requires an infinite number of CQI samples. In our work, we provide analytical results which can be used to evaluate system performance for any finite value N_k.

As discussed in Section 4.3.3, there are two options for dividing the interval between rate r_i and r_{i+1}. The analysis for both options is similar and so we only provide analytical results for the case where the interval is divided into equally spaced intervals in the rate domain. The interval between rate r_{i_k} and r_{i_k+1} is divided into L_{i_k} equally spaced intervals ($r_{i_k} = r_{i_k,0}, r_{i_k,1}, \ldots, r_{i_k,L_{i_k}} = r_{i_k+1}$). When a user CQI is associated with a rate region $r_{i_k,j}$, the channel can support transmission rate $r_{i_k,j}$. However, as the transmission rates supported are discrete and only M, the system is assumed to only transmit at rate $r_{i_k,0} = r_{i_k}$.

For the analysis, an important step is obtaining an expression for the probability that user k is selected when it has the rate $r_{i_k,j}$. We now develop this expression. Among $(N_k - 1)$ CQI samples, we denote by $\hat{m}_{i_k,j}$ the number of collected samples that support rates smaller than or equal to $r_{i_k,j}$, and by $(m_{i_k,j} - 1)$ the number of samples, other than the most recent sample, that support rate $r_{i_k,j}$. Then, the number of CQI samples, excluding the current instantaneous sample, that support a rate smaller than $r_{i_k,j+1}$ is $\hat{m}_{i_k,j+1} = \hat{m}_{i_k,j} + m_{i_k,j}$. Using the binomial distribution [76], the probability there are $\hat{m}_{i_k,j+1} = \hat{m}_{i_k,j} + m_{i_k,j} - 1$ samples that support rate smaller than $r_{i_k,j+1}$ is [53, 76]

$$\Pr\{\hat{m}_{i_k,j+1} = \hat{m}_{i_k,j} + m_{i_k,j} - 1\} = \binom{N_k-1}{\hat{m}_{i_k,j}+m_{i_k,j}-1} F_{R_k}(r_{i_k,j+1})^{\hat{m}_{i_k,j}+m_{i_k,j}-1} \times (1 - F_{R_k}(r_{i_k,j+1}))^{N_k-\hat{m}_{i_k,j}-m_{i_k,j}}. \quad (4.7)$$

Similarly, among these $\hat{m}_{i_k,j} + m_{i_k,j} - 1$ samples, the probability a sample supports a rate smaller than $r_{i_k,j}$ is $\frac{F_{R_k}(r_{i_k,j})}{F_{R_k}(r_{i_k,j+1})}$. We consider the event that there are $\hat{m}_{i_k,j}$ samples that support a rate smaller than $r_{i_k,j}$ and $m_{i_k,j} - 1$ samples with achievable rate between $[r_{i_k,j}, r_{i_k,j+1})$. Given $\hat{m}_{i_k,j+1}$, the probability of this event is given by

$$\Pr\{\hat{m}_{i_k,j}, m_{i_k,j}|\hat{m}_{i_k,j+1}\} = \binom{\hat{m}_{i_k,j}+m_{i_k,j}-1}{m_{i_k,j}-1} (F_{R_k}(r_{i_k,j+1}) - F_{R_k}(r_{i_k,j}))^{m_{i_k,j}-1} \times F_{R_k}(r_{i_k,j})^{\hat{m}_{i_k,j}} F_{R_k}(r_{i_k,j+1})^{1-\hat{m}_{i_k,j}-m_{i_k,j}}. \quad (4.8)$$

Then, by combining (4.7) and (4.8) and using Bayes rule

$$\Pr\{\hat{m}_{i_k,j}, m_{i_k,j}\} = \binom{N_k-1}{\hat{m}_{i_k,j}+m_{i_k,j}-1} \binom{\hat{m}_{i_k,j}+m_{i_k,j}-1}{m_{i_k,j}-1} F_{R_k}(r_{i_k,j})^{\hat{m}_{i_k,j}} (F_{R_k}(r_{i_k,j+1}) - F_{R_k}(r_{i_k,j}))^{m_{i_k,j}-1} (1 - F_{R_k}(r_{i_k,j+1}))^{N_k-\hat{m}_{i_k,j}-m_{i_k,j}}. \quad (4.9)$$

Given a set of $\hat{m}_{i_k,j}$ and $m_{i_k,j} - 1$ samples, the scheduler generates a random variable $U_{k,i_k,j}$ which is uniformly distributed in $\left[\frac{\hat{m}_{i_k,j}}{N_k}, \frac{\hat{m}_{i_k,j}+m_{i_k,j}}{N_k}\right]$. The probability user k is selected given this condition is [58]

$$\Pr\{k^* = k|\hat{m}_{i_k,j}, m_{i_k,j}\} = \frac{w_k N_k}{m_{i_k,j}} \left(\left(\frac{\hat{m}_{i_k,j} + m_{i_k,j}}{N_k} \right)^{\frac{1}{w_k}} - \left(\frac{\hat{m}_{i_k,j}}{N_k} \right)^{\frac{1}{w_k}} \right). \quad (4.10)$$

With all these components in place, finally we have

$$\Pr\{k^* = k|i_k, j\} = \sum_{\Pi(\hat{m}_{i_k,j}, m_{i_k,j})} \Pr\{k^* = k|\hat{m}_{i_k,j}, m_{i_k,j}\} \Pr\{\hat{m}_{i_k,j}, m_{i_k,j}\} \quad (4.11)$$

where $\Pi(\hat{m}_{i_k,j}, m_{i_k,j})$ denotes all possible combinations of $\hat{m}_{i_k,j}$ and $m_{i_k,j}$. Equation (4.11) can be computed utilizing Equations (4.9) and (4.10)

Another probability required before we can evaluate system performance is the probability that a user k has CQI corresponding to the rate index i_k, and rate sub-region j. This is given by

$$\Pr\{i_k, j\} = F_{R_k}(r_{i_k,j+1}) - F_{R_k}(r_{i_k,j}). \tag{4.12}$$

System throughput in the CSwQ

When the channel of a user k has CQI represented by the index $\{i_k, j\}$, the BS chooses a rate r_{i_k} for the link. The system throughput is given by

$$R = \sum_{k=1}^{K} \sum_{i_k=0}^{M-1} \sum_{j=0}^{L_{i_k}-1} r_{i_k,j} \Pr\{k^* = k | i_k, j\} \Pr\{i_k, j\}. \tag{4.13}$$

Substituting (4.11) and (4.12) into (4.13), we have the system throughput

$$\begin{aligned} R = \sum_{k=1}^{K} w_k \sum_{i_k=0}^{M-1} r_{i_k} \sum_{j=0}^{L_{i_k}-1} \sum_{\Pi(\hat{m}_{i_k,j}, m_{i_k,j})} \frac{N_k}{m_{i_k,j}} \left(\left(\frac{\hat{m}_{i_k,j} + m_{i_k,j}}{N_k} \right)^{\frac{1}{w_k}} - \left(\frac{\hat{m}_{i_k,j}}{N_k} \right)^{\frac{1}{w_k}} \right) \\ \times \binom{N_k - 1}{\hat{m}_{i_k,j} + m_{i_k,j} - 1} \binom{\hat{m}_{i_k,j} + m_{i_k,j} - 1}{m_{i_k,j} - 1} F_{R_k}(r_{i_k,j})^{\hat{m}_{i_k,j}} \\ (F_{R_k}(r_{i_k,j+1}) - F_{R_k}(r_{i_k,j}))^{m_{i_k,j}} (1 - F_{R_k}(r_{i_k,j+1}))^{N_k - \hat{m}_{i_k,j} - m_{i_k,j}}. \end{aligned} \tag{4.14}$$

Power savings in the CSwQ

In the proposed method, when a link is selected it is likely that the channel is able to support a rate higher than the M rates supported by the system. Hence, the BS can potentially back off its transmit SNR to $\hat{\rho}$ and be able to support the rate of transmission. For this purpose, we need to find the smallest value of $\hat{\rho}$ which supports rate r_{i_k}.

Specifically, with transmit SNR ρ, user k can support rate $r_{i_k,j} = \log_2(1 + \rho x_k)$ where x_k is the channel CQI. Then, $x_k \geq \frac{2^{r_{i_k,j}} - 1}{\rho}$. However, since the actual rate transmitted is $r_{i_k,0}$, where $r_{i_k,0} = \log_2(1 + \hat{\rho} x_k)$, the transmit SNR $\hat{\rho}$ can be lowered and given by

$$\hat{\rho} \geq \frac{2^{r_{i_k,0}} - 1}{2^{r_{i_k,j}} - 1} \rho, \tag{4.15}$$

where $r_{i_k,j} = r_{i_k,0} + j\Delta r_{i_k}$, with $\Delta r_{i_k} = \frac{1}{L_{i_k}}(r_{i_k+1} - r_{i_k})$.

The average power saving can be calculated similar to (4.13)

$$E\{\rho - \hat{\rho}\} = \sum_{k=1}^{K} \sum_{i_k=0}^{M-1} \sum_{j=0}^{L_{i_k}-1} \rho \left(1 - \frac{2^{r_{i_k,0}} - 1}{2^{r_{i_k,j}} - 1}\right) \Pr\{k^* = k | i_k, j\} \Pr\{i_k, j\}. \qquad (4.16)$$

Substituting (4.11) and (4.12) into (4.16), the average power saved because of the additional feedback resource employed to get finer quantization regions is given by

$$\begin{aligned} E\{\rho - \hat{\rho}\} = &\sum_{k=1}^{K} w_k \sum_{i_k=0}^{M-1} \sum_{j=0}^{L_{i_k}-1} \rho \left(1 - \frac{2^{r_{i_k,0}} - 1}{2^{r_{i_k,j}} - 1}\right) \sum_{\Pi(\hat{m}_{i_k,j}, m_{i_k,j})} \frac{N_k}{m_{i_k,j}} \\ &\times \left(\left(\frac{\hat{m}_{i_k,j} + m_{i_k,j}}{N_k} \right)^{\frac{1}{w_k}} - \left(\frac{\hat{m}_{i_k,j}}{N_k} \right)^{\frac{1}{w_k}} \right) \binom{N_k - 1}{\hat{m}_{i_k,j} + m_{i_k,j} - 1} \binom{\hat{m}_{i_k,j} + m_{i_k,j} - 1}{m_{i_k,j} - 1} \\ &\times F_{R_k}(r_{i_k,j})^{\hat{m}_{i_k,j}} (F_{R_k}(r_{i_k,j+1}) - F_{R_k}(r_{i_k,j}))^{m_{i_k,j}} (1 - F_{R_k}(r_{i_k,j+1}))^{N_k - \hat{m}_{i_k,j} - m_{i_k,j}}. \end{aligned} \qquad (4.17)$$

To evaluate the achievable rate of each user k, we simply substitute the appropriate channel distribution function into (4.17). The parameters of how to cluster each of the rate regions are based on the observed data as described in the Alg. 4.

Proposition 4. *When the number of collected CQI goes to infinity, the saving in power consumption of the CSwQ is*

$$E\{\rho - \hat{\rho}\} = w_k \sum_{k=1}^{K} \sum_{i_k=0}^{M-1} \sum_{j=0}^{L_{i_k}-1} \rho \left(1 - \frac{2^{r_{i_k,0}} - 1}{2^{r_{i_k,j}} - 1}\right) \left(F_{R_k}(r_{i_k,j+1})^{\frac{1}{w_k}} - F_{R_k}(r_{i_k,j})^{\frac{1}{w_k}} \right), \qquad (4.18)$$

and the average system throughput [1] is

$$R = w_k \sum_{k=1}^{K} \sum_{i_k=0}^{M-1} r_{i_k} \left(F_{R_k}(r_{i_k+1})^{\frac{1}{w_k}} - F_{R_k}(r_{i_k})^{\frac{1}{w_k}} \right). \qquad (4.19)$$

Proof. From (4.17), when the number of observed samples N_k gets larger, the fractions converge to the CDF value, e.g. $\frac{\hat{m}_{i_k,j}}{N_k} \to F_{R_k}(r_{i_k,j})$. Then, given the feedback index $\{i_k, j\}$ of user k, the random variable $U_{k,i_k,j}$ is uniformly distributed in $\left[\frac{\hat{m}_{i_k,j}}{N_k}, \frac{\hat{m}_{i_k,j}+m_{i_k,j}}{N_k}\right] \equiv [F_{R_k}(r_{i_k,j}), F_{R_k}(r_{i_k,j+1})]$. The probability that user k is selected is given by [58]

$$\Pr\{k^* = k|i_k, j\} = w_k \frac{F_{R_k}(r_{i_k,j+1})^{\frac{1}{w_k}} - F_{R_k}(r_{i_k,j})^{\frac{1}{w_k}}}{F_{R_k}(r_{i_k,j+1}) - F_{R_k}(r_{i_k,j})}. \tag{4.20}$$

Combine (4.12), (4.15), and (4.20) we have (4.18). Similarly, we can obtain the average system throughput which is equal to the throughput of extended CS [1] and AeCS.

□

Note that the above analysis is general and no specific channel model is assumed. Each channel distribution function $F_{R_k}(\cdot)$ represents the channel characteristics of a specific user which incorporates channel pathloss and fading. The system performance for different channel distributions can be easily studied. For instance, if we assume all users are under Rayleigh fading with SNR distribution $F_{X_k}(x) = 1 - e^{-\frac{x}{\rho_k}}, x \geq 0$, the channel distribution can be expressed as

$$F_{R_k}(r) = 1 - e^{-\frac{2^r-1}{\rho_k}}, r \geq 0 \tag{4.21}$$

where throughput r is a function of the received SNR x, i.e. $r = \log_2(1+\rho_k x)$. The received SNR of user k is $\rho_k = c_k\rho$ and is a function of the transmit SNR ρ and the pathloss c_k. This distribution function is then substituted into (4.14) and (4.17) to evaluate system throughput and power savings in our proposed scheduling scheme.

4.4 Simulation results

We consider a multiuser downlink system with $K = 10$ users, where a user k is assigned weight $w_k = ak$ with the constant a chosen such that $\sum_{k=1}^{K} w_k = 1$. The channel is assumed to be experiencing Rayleigh fading. The exponential pathloss model is used where the passloss of user k is set $Ce^{-\lambda k}$ with parameters $\lambda = 0.1$

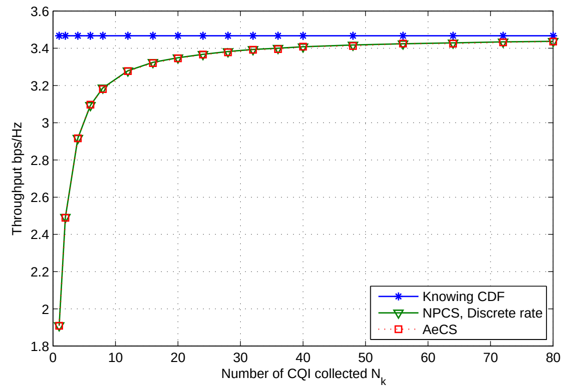


Figure 4.1: Average throughput of CDF scheduling methods under Rayleigh fading with explicit feedback.

and C is chosen so that $\sum_{k=1}^{K} Ce^{-\lambda k} = K$. The transmit SNR is set at $\rho = 10$ dB. The setup used in our simulations is chosen to reflect the diversity in weights and pathloss of the users as well as to facilitate a systematic and controlled study. These weights and pathloss value are then used for all the experiments for the sake of fair comparison between the investigated scheduling methods. Other scenarios have been simulated with similar conclusions.

4.4.1 Explicit feedback

Though we develop techniques for a system with limited feedback, system performance when explicit feedback is available at the BS is considered at first. To be specific, the BS has exact CQI values and applies the scheduling methods together with the power back off scheme. The purpose of this experiment is to investigate the sole effect of finite number of transmission rates on overall system performance. Then, the impact of limited feedback is introduced and more clearly demonstrated in the other experiments.

In Fig. 4.1, the system throughput as a function of the number of collected CQI samples is investigated. In this experiment, the maximum transmit SNR is fixed at $\rho = 10$ dB. The BS uses the following set of transmission rates $[0, 1, 2, 4, 6, 8]$ bps/Hz. When the system has more CQI samples (N_k), the BS can select a user for service more precisely and so the system has higher throughput.

With discrete number of rates, NPCS [58] and AeCS achieve the similar system throughput. When N_k gets larger, system throughput achieved by NPCS or the extended CS method approach the throughput achieved when the CDF is perfectly known.

In Fig. 4.2, average system transmit power as a function of the number of collected CQI samples is shown. From an energy perspective, knowing the CDF enables better exploitation of multiuser diversity and thereby results in reduced transmit power. NPCS [58] uses less transmit power than AeCS to support the same average system throughput. Given the two methods are both able to back off the transmit power, the savings in transmit power is about 1 dB as shown in Fig. 4.2. Alternately, using the same average transmit power, NPCS achieves higher system throughput. When the number of collected samples N_k increases, as shown in Fig. 4.1 and 4.2, NPCS converges rapidly to CDF scheduling with perfect CDF knowledge in terms of both average system throughput and transmit power.

4.4.2 Limited feedback

We now examine the performance of the methods developed to deal with limited feedback and finite number of transmission rates.

In Fig. 4.3, the average system power consumption as a function of the number of feedback bits is shown. As can be seen from the figure, the modified NPCS and the CSwQ methods use energy very efficiently. Similar to the case with explicit feedback, the average consumed power is about 1 dB less than that in the AeCS when the power saving mechanism is applied. Moreover, when the system assigns more feedback bits to encode each CQI sample, the power consumption of these methods converge very rapidly to the power consumption when the CDF is perfectly known.

In Fig. 4.4, system throughput as a function of number of feedback bits is shown. As can be seen from the figure, as long as the feedback bits are adequate to encode all the transmission rate regions, the throughput achieved by the AeCS and the CSwQ remains unchanged and no longer depend on the number of feedback bits. Moreover, it can be seen that the AeCS and the CSwQ have the same throughput because the users in these methods have the same probability to be selected for any transmission rate. The weakness of the modified NPCS algorithm discussed before, namely slightly lower system throughput, is verified. The loss is due to the mismatches between the boundaries used for user selection and the rate boundaries. If one user selection region includes two rate regions, this will cause an increase in the probability that a user with lower rate is selected. When the system has more feedback resource (bits per CQI), as expected, the loss diminishes and the system throughput using the modified NPCS method gradually approach that of AeCS.

The numerical results show that the developed CSwQ inherits the virtues of modified NPCS and AeCS without their respective limitations. First, as can be seen from Fig. 4.4, it has eliminated the loss in throughput suffered by modified NPCS by setting the feedback intervals not to overlap with more than one transmission rate regions. As can be seen from Figure 4.3, it has successfully exploited the additional feedback resource, by dividing the rate regions into smaller regions effectively, to enable power savings in contrast to AeCS. From Fig. 4.4, the CSwQ has similar performance as the AeCS in term of system throughput. However, it uses less power than AeCS as it selects users more effectively.

We now validate the analytical results. The throughput (Equation (4.14)) and power savings (Equation (4.17)) of the CSwQ method are verified in Fig. 4.5 and 4.6. In Fig. 4.5, the achievable throughput of user k as a function of number of collected CQI N_k is shown. In this experiment, it is assumed that the system has 3 bits to encode each CQI sample and there are only 6 possible transmission rates. Hence, the additional freedom in quantization is used to further subdivide the transmission rate regions $[2, 4]$ and $[4, 6]$ into two intervals for power savings purpose. From the figures it can be seen that the analytical and simulated throughput and power saving results agree well. The analytical results are not asymptotic results, and show analytically how the actual throughput depends on the number of sample N_k. In Fig. 4.6, the consumed power of a user k as a function of number of CQI samples N_k is shown. In these figures, the convergence rate is observed to be quite rapid. The throughput and power savings derived in (4.14) and (4.17) quickly converge to the corresponding asymptotic values in Proposition 4.

Through the experiments, we observe that there is not only overall system throughput but also always a proportional improvement for each individual user. The setting, e.g weights and pathloss of users, only affects the magnitude of the improvement.

4.5 Conclusions

In this work, simple CDF scheduling methods are developed which can work without a priori knowledge of the channel distribution and can be employed in systems that support discrete number of transmission rates and have limited feedback resources. The developed methods are AeCS, modified NPCS and the CSwQ. Among these methods, the CSwQ is shown to incorporate the virtues of the other two methods without inheriting their drawbacks. As a result, the CSwQ achieves the highest system throughput while effectively saving transmit power. System throughput of the CSwQ is then analyzed and shown to converge rapidly to the performance with increasing number of CQI samples. Through simulations, the developed methods are shown to converge rapidly to the idealistic case of systems with perfect channel CDF, achieve the highest throughput for finite rate systems, and effectively save transmit power.

The text of this chapter, in full, is a reprint of the paper [77], Anh H. Nguyen, and Bhaskar D. Rao CDF scheduling algorithms for finite rate multiuser systems with limited feedback, that has been submitted for publication on IEEE Transaction on Wireless Communications. The dissertation author is the primary researcher and author, and the co-author listed in this publication guided and supervised the research constitutes this chapter.

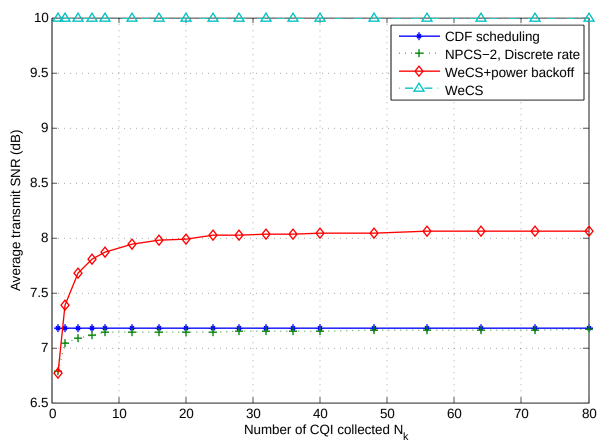


Figure 4.2: Average power assumption of CDF scheduling methods with finite rate transmission under Rayleigh fading with explicit feedback.

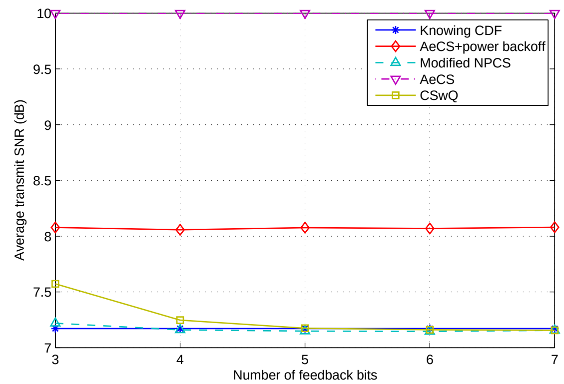


Figure 4.3: Average power consumption of CDF scheduling methods under Rayleigh fading with B bits of quantized feedback.

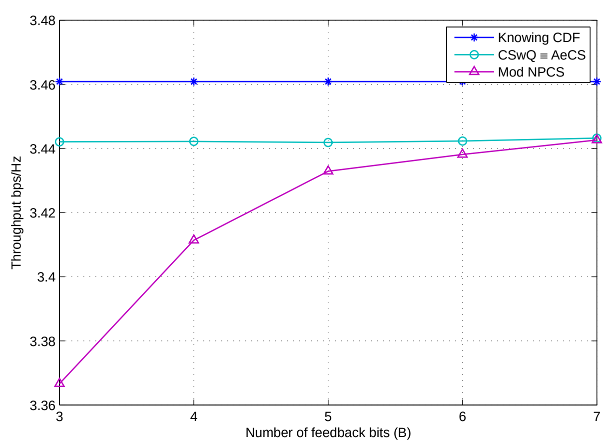


Figure 4.4: Average throughput of CDF scheduling methods under Rayleigh fading with B bits of quantized feedback.

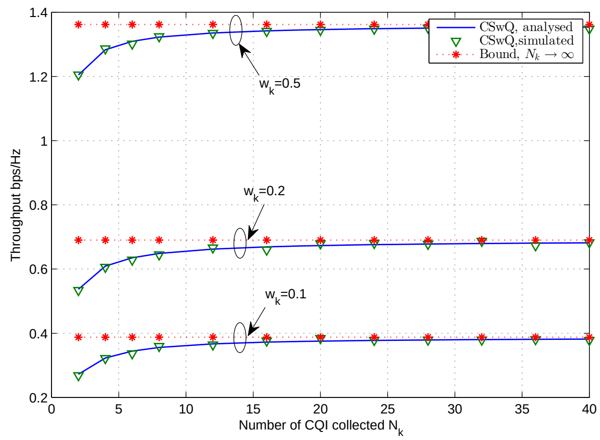


Figure 4.5: Analytical results of achievable throughput of a user k using the CSwQ scheme under Rayleigh fading.

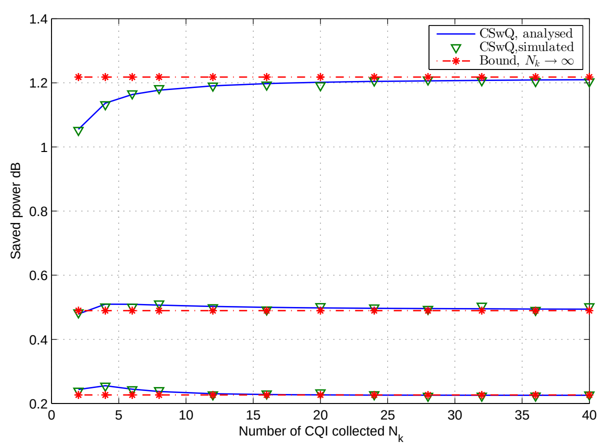


Figure 4.6: Power saving for user k using the CSwQ scheme under Rayleigh fading.

Chapter 5

Weighted CDF-based Scheduling for an OFDMA Relay Downlink with Partial Feedback

The performance of partial feedback OFDMA systems which have users served by a macro base-station (BS) or by users served by both macro BS and relays is analyzed in this paper. To reduce feedback, each user feeds back only the best M channel quality information (CQI) among the total number of resource blocks. A weighted cumulative distribution function (CDF) based scheduling is utilized to help both ensure fairness among users and exploit multiuser diversity. A general analysis is provided which is first applied for the system with only macro BS and then extended to the system with both a macro BS and relays. Herein, the distribution and the average throughput are expressed in polynomial forms which are convenient for further study. Moreover, it is shown that the allocation probability for the users can be met by appropriate choice of weights which are unique and can be found efficiently. When adding relays to the system, the different time scale on the two hops of the relay are observed to cause resource starvation for users. A modified version of the weighted CDF based scheduling is developed to maintain fairness in a short term manner. The system is further enhanced by exploiting the imbalance in the signal to noise ratio (SNR) on the two hops of a relay to save transmit power. Finally, experiments are conducted to develop insight and verify the analytical results.

5.1 Introduction

Wireless systems have the challenge of serving the needs of users using various services with different rates and quality of service requirements. Besides, mobile users in the system can have arbitrary locations in the cell which leads to a diversity of pathloss and fading conditions as well as channel statistics. In addition, a period that a user is in need of resource depends on the characteristics of his data application which is typically independent of the user's channel condition. The variability of channel quality of users in both time and space makes resource allocation for users based on their demands a challenging task. In particular, exploiting multiuser diversity while ensuring different service requirements (fairness) across users poses a significant challenge. To help users at the cell edge, who frequently have poor channel conditions due to large pathloss and severe fading, relays have been suggested to pass on the signal from the macro base station (BS) to the mobile user. A relay typically has larger demand on resource as it serves a group of cell edge users and naturally gives rise to the diversity of rate requirement scenario. Moreover, its channel might have a different fading time scale (channel coherence time) further complicating the allocation strategy. Motivated by these issues, in this work we consider a resource allocation problem which concurrently provides fairness for users and efficiently exploits multiuser diversity in the context of partial channel information and diversity of rate requirements. Then, the allocation strategy is altered to be used in a system with the involvement of relays. An approach is proposed to keep guaranteeing fairness among users in the presence of the different fading time scales on the links of the two hops of relays.

It is well known that fully exploiting the diversity in the channels of users, also known as multiuser diversity, helps to achieve a higher system sum rate. In the uplink of single cell multiuser systems, the maximum capacity is achieved by selecting the user with the best channel condition [2]. This exploitation of multiuser diversity can be incorporated with spatial diversity in MIMO systems as shown in [78]. For wideband OFDMA systems, the capacity of the multiuser system [79] is obtained by optimizing power allocation on each sub-carriers as well as among users. A greedy approach is used in [80] to find the best set of users and is shown to offer a higher sum rate for wireless systems. Herein, it is desirable to capture multiuser diversity by selecting users whenever their channels are in good condition.

Besides, as fairness among users is also of importance in wireless systems, obtaining the highest system throughput while guaranteeing fairness is desirable. These objectives are somehow conflicting because a throughput centric resource allocation strategy leads to the fact that some users with good channel are always allocated the resource while the others are starved of resources. This situation is typically caused by both near-far phenomena and different fading statistics. In wireless communications, guaranteeing service as well as maintaining the fairness among users is hence one of the major concerns. In [81] system performance using opportunistic beamforming is investigated. To support fairness, round robin scheduling and several different schedulings methods are compared in [29]. Though round robin can support fairness, it can not exploit multiuser diversity. A commonly used method to both guarantee fairness and exploit multiuser diversity is proportional fair scheduling [82]. There are several flavors of fair scheduling algorithms, and a comparison in term of both rate fairness and delay minimization can be found in [26]. The work in [45] investigates performance of different variations of proportional fair scheduling in comparison with the round robin method. Though proportional fair can efficiently guarantee fairness among users, analytical tractability of its performance has proven difficult. Using proportional fair, the system performance or particularly system throughput can not be easily expressed in a convenient form. we do not have a convenient performance metric which could be used to further analyze or optimize system performance if additional system parameters are taken into account. In this work we consider weighted CDF-based scheduling [21]. This is motivated by the fact that CDF based scheduling also has the ability to exploit multi-user diversity, ensure fairness and support a diversity of rates. In [48], the CDF-based scheduling is leveraged in a general multicell network to guarantee scheduling fairness and simultaneously obtain multiuser diversity gain when different users experience diverse intercell interference. In [50], the CDF-based scheduling is analytically studied in the random beamforming framework, and the notion of individual sum rate and individual scaling laws are proposed to characterize the performance under this scheduling policy. Equal weights across users are assumed in both [48] and [50]. In this paper, the challenge is to make use of generic weighted CDF based scheduling in a system with partial feedback and with the participation of relays, where the distribution of the channels on the two hops is significantly different and also have a vastly differing channel coherence time.

To concurrently exploit mulituser diversity and guarantee fairness among users, channel information is needed at the base station [23, 82]. If the total feedback resource is strictly constrained multiuser diversity can not be exploited [65]. In a frequency division duplex system, the feedback requirement can be overburden as the frequency dimension is added. A wideband system typically requires more feedback to capture both multiuser diversity and frequency diversity imposing practical constraints [24]. Specifically, a large number of resource blocks in wideband multi-user MIMO OFDM systems will require a huge amount of feedback [23], and it is of utmost importance to reduce feedback requirement as well as find scheduling techniques which best utilize such limited feedback information. Interestingly, multiuser diversity can still be exploited efficiently with partial feedback [22]. Various types of feedback techniques are discussed in [83]. In [71] a threshold based feedback scheme is considered while in [48, 84] a partial feedback scheme is investigated where only the best M CQIs among the overall N resource blocks are fed back. The work in [85] considers grouping highly correlated resource blocks to reduce feedback. In our work, we analyze an OFDMA downlink with best M CQI feedback.

Having examined the problem in a traditional macro BS setting, we are also interested in considering the incorporation of relays into wireless systems. As is well known, the uniformity of service requirement raises a stringent requirement on system coverage and capacity, especially in the cell edge region. Among the complementary technology, the use of relays helps increase network coverage and throughput. Furthermore, multiuser diversity can still be exploited in a system with relays [86]. In [87, 88] multiple single antenna relays are deployed to aid the communications from the BS to users. In [89], performance of Amplify-and-Forward (AF) and Decode-and-Forward (DF) relays are compared. In the dual-hop relay system [90], the relays aid the transmission when the user can not decode all information in the superposition code. In our work, DF is chosen to aid scheduling and it enables us to handle noise more easily as the relay does not amplify noise in the signal. In [91], an overview and performance comparison of different relaying techniques in a wireless network are provided. Meanwhile, different relaying strategies for both Decode-and-Forward and Amplify-and-Forward relays using Space-Time signaling are discussed and evaluated in [92]. Spectral efficiency of protocols including half-duplex, two-way and two-path relaying are also discussed [93]. Being extended to MIMO system in [94], tight lower and upper bounds for the capacity of MIMO Relay channels under Rayleigh fading are derived. Then, the optimal power allocation is found for minimizing outage for multi-hop relayed transmissions over Rayleigh fading channels in [95]. This work addresses the challenge in scheduling that arises due to the two hops of the relay network. In particular, the differing channel coherence times of the two hops possess new challenges in the fairness and multiuser diversity maximization tradeoff.

The outline and main contributions of the paper are as follows. In Section II the system model is described, and in Section III we conduct analysis to investigate the performance of the system. Firstly, we apply and analyze weighted CDF-based scheduling for OFDMA system with only macro users, i.e. users served by the macro BS. Partial feedback is used where each of the users feeds back only the best M channels among all the resource blocks. By using the weighted CDF-based scheduling method, probability of allocation to users can be effectively controlled in a long term manner. The analysis[1] provides the CDF of the SNR of the selected users and the average system throughput in the form of polynomials which are convenient for further evaluation. We also propose an algorithm which is able to effectively find appropriate weights for the users commensurate with their priority. Secondly, in Section IV, the scheduling method is applied to an OFDMA system with relays to manage the resource in the two hop network. We consider both options for the channel coherence time, fast Rayleigh fading and slow log-normal fading [96], for the channels between the BS and the relay. With the slow fading assumption, new challenges arise in the scheduling process. The slowly varying BS-relay links can cause users served by a relay to be starved of resources for a long time when the link undergoes severe fading condition. We propose a novel scheme to interpolate the CDF of all the relay links so that we are still able to guarantee short term fairness for all users. Lastly, the imbalance in the quality of the two hops of a relay is utilized to enhance system performance by reducing the transmit power of either BS or the relay. Simulation results are used to validate the analysis and illustrate the performance of the proposed techniques.

[1]A preliminary version of this work was presented in [49]

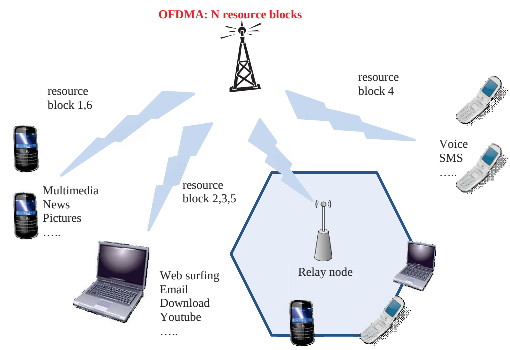


Figure 5.1: An OFDMA system with $N = 10$ resource blocks, has different macro users, and 1 relay serving 3 users.

5.2 System model

We consider a multiuser downlink single-input single-output (SISO) OFDMA system with a macro base station equipped with a single antenna and K users. Each user k is assigned a certain priority reflected by a weight w_k as depicted in Fig. 5.1. All the equipments in the network have a single antenna. The formulation can accommodate two network models. In one model, the system has K users communicating directly with the BS, while in the other model some of the users are connected through relays. All these diverse users compete for resources. Intuitively, each group of users communicating through a relay can be considered as a single "user" with a certain priority. We will deal with the scenario containing relays in more detail later. For now, we assume all the users are macro users, i.e. directly served by the macro base station.

In the OFDMA system, the total bandwidth is divided into N independent resource blocks. On each of the resource block, the system is a multiuser downlink SISO system where the channels for a single user is flat on each resource block, and i.i.d. across resource blocks. This model is a simplification of reality, but forms a useful basis for analysis and comparative study. We denote $h_{k,r}$ as the channel from the BS to user k on resource block r. The received signal at the receiver k is

$$y_{k,r} = \sqrt{P} h_{k,r} s_{k,r} + n_{k,r}, \tag{5.1}$$

where $y_{k,r}$ is the received signal, $s_{k,r}$ is the transmitted signal, P is the average transmit power, and $n_{k,r} \sim CN(0, \sigma_w^2)$ is the independent noise at user k on resource block r. Because channels on every resource block are assumed statistically identical, from now on, all the analysis is conducted on a single resource block r.

5.2.1 Partial feedback

Because obtaining full feedback in such a spatial-frequency large scale network is prohibitive in feedback resources, we consider partial feedback where each user feeds back CQI information of the best M resource blocks among it's N resource blocks with $M < N$[48, 84]. Due to this partial feedback scheme, on a certain resource block there is only a subset of users who provide feedback. On resource block r, we denote S_r, with cardinality $|S_r|$, the set of users who feed back. In this case, the cardinality $|S_r|$ is a random variable. Knowing that S_r can consist of any combination of users, $\pi(S_r)$ is used to denote all the possible combinations.

5.2.2 CDF based scheduling

To control the assignment probability, weighted CDF-based scheduling [21] is used where a user is selected if it achieves high but not likely to obtain higher throughput. In the system, each user k is assigned a weight w_k which represents its priority. The weight w_k is also the proportion of the resource allocated by the network to the user. For example, if the total network resource is 10 Mbps which corresponded to 100% and user k is allocated 10% airtime which might be in the order of 1 Mbps, then the weight for this user is set $w_k = \frac{1}{10} = 0.1$. The weight w_k is preassigned for all the users $k = 1, \ldots, K$ such that $\sum_{k=1}^{K} w_k = 1$.

On each resource block, the BS upon receiving several CQI from some of the users, assumed to be $x_1, \ldots, x_n$. The BS utilizes the corresponding CDF of these CQI to evaluate a service metric and selects the user k^* with the highest value. The CDF scheduler selects the user to be served as follows:

$$k^* = \arg\max_k F_{X_k}(x_k)^{\frac{1}{w_k}}, \tag{5.2}$$

where $F_{X_k}(.)$ is the CDF of the CQI of user k.

5.3 OFDMA systems with prioritized users

In order to analyze system performance, we want to find the distribution of users' SNR when they are selected and use the found distribution to calculate system throughput. Similar to [48, 84], to make the analysis easy to follow, the procedure is divided into the following steps.

- **CQI:** Each receiver k knows its own CQI on every resource block $r = 1, \ldots, N$. The CQI is represented by the variable $Z_{k,r} = |h_{k,r}|^2$ and has distribution $F_{Z_k}(\cdot)$ which can be different for each user and for different system models.

- **Best M feedback:** Each user k then selects the M best CQI among the total N resource blocks and feeds back this information to the base station (BS). The BS receives these CQI and sees the fed back CQI as a random variable $Y_{k,r}$ with distribution $F_{Y_k}(\cdot)$.

- **Scheduling:** On each resource block, the BS has CQI from a random subset of users and it selects the user with the best CQI based on the proposed weighted CDF scheduling strategy. The SNR of the selected user is X_r with distribution $F_{X|S_r}(\cdot)$, where S_r denotes the subset of users who provided feedback on that resource block.

- **System throughput:** Based on the distribution of the conditional SNR, the CDF and average value of system throughput are found by taking expectation over all feedback possibilities.

5.3.1 Analysis of a General OFDMA system with best M feedback and user priority

Now, we use the steps described above to perform the analysis.

1. **Step 1: Find** $F_{Z_k}(\cdot)$ As described earlier, $Z_{k,r} = |h_{k,r}|^2$ represents the CQI from the BS to user k on resource block r. Based on the specific system model and type of fading, we can determine the distribution function $F_{Z_k}(\cdot)$. An example is a exponential distribution. For now, the analysis is kept general and done without assuming a specific form. The expression obtained will later be specialized for specific channel models.

2. **Step 2: Find** $F_{Y_k}(\cdot)$ We denote $Y_{k,r}$ the SNR corresponding to received CQI at the transmitter for user k on block r through feedback. The transmitter views the SNR as any of the best M CQI values multiplied by $\rho = P/\sigma_w^2$ [84], where P, σ_w and ρ are power, noise variance and SNR on each resource block as in (5.1). It can be shown to have the distribution given by [84]

$$F_{Y_k}(x) = \sum_{m=0}^{M-1} e_1(N, M, m) \left\{ F_{Z_k}\left(\frac{x}{\rho}\right) \right\}^{N-m}, \tag{5.3}$$

with $e_1(m) = \sum_{i=m}^{M-1} \frac{M-i}{M} \binom{N}{i} \binom{i}{m} (-1)^{i-m}$. For notational convenience, we will use $e_1(m)$ instead of $e_1(N, M, m)$.

3. **Step 3: Scheduling** The probability that a certain user provides feedback on a resource block is $\frac{M}{N}$. Using binomial distribution, the probability that a certain set of users S_r provide feedback is

$$\Pr\{S_r\} = \left(\frac{M}{N}\right)^{|S_r|} \left(1 - \frac{M}{N}\right)^{K-|S_r|}. \tag{5.4}$$

Let k_r^* denote the selected user for transmission on block r. The probability user k is selected, given it is in the set S_r, is $\Pr\{k_r^* = k | k \in S_r\} = \frac{w_k}{\sum_{j \in S_r} w_j}$.

Then, the total probability user k is selected, is

$$\Pr\{k_r^* = k\} = \sum_{\pi(S_r|k\in S_r)} \frac{w_k}{\sum_{j\in S_r} w_j} \left(\frac{M}{N}\right)^{|S_r|} \left(1-\frac{M}{N}\right)^{K-|S_r|}, \qquad (5.5)$$

where $\pi(S_r)$ denote all possible set of users S_r.

We denote $U_{k,r}$, as a random variable obtained from the mapping $u_{k,r} = F_{Y_k}(y_k)$. Among all the users in S_r, we select the user with highest value of $U_{k,r}^{\frac{1}{w_k}}$. The conditional CDF of X_r, the SNR of the selected user at the transmitter can be derived to be:

$$F_{X|k_r^*=k,S_r}(x) = \prod_{j\in S_r} \Pr\{u_{j,r}^{\frac{1}{w_j}} \leq u_{k,r}^{\frac{1}{w_k}}\} = \{F_{Y_k}(x)\}^{\frac{\sum_{j\in S_r} w_j}{w_k}}. \qquad (5.6)$$

4. **Step 4: System throughput** In this step, we derive the CDF of the system SNR as well as the average system throughput by taking the expectation of the conditional CDF over the conditioning variable, the set of users who provide feedback.

Theorem 5. *In a single antenna, K user, N resource block, M best feedback OFDMA system employing weighted CDF based scheduling with weight w_k for user k, the CDF of the SNR of the selected user is given by*

$$F_{X_r}(x) = \sum_{k=1}^{K} \sum_{\pi(S_r|k\in S_r)} \frac{w_k}{\sum_{j\in S_r} w_j} \left(\frac{M}{N}\right)^{|S_r|} \left(1-\frac{M}{N}\right)^{K-|S_r|} \times \sum_{t=1}^{\infty} e_3(\alpha_k, t) \sum_{m=0}^{t(M-1)} e_2(t,m) F_{Z_k}\left(\frac{x}{\rho}\right)^{Nt-m}, \qquad (5.7)$$

where $e_3(m) = \sum_{i=t}^{\infty} \binom{\alpha_k}{i} \binom{i}{t} (-1)^{i-t}$, with $\alpha_k = \frac{\sum_{j\in S_r} w_j}{w_k}$ and $e_2(0) = e_1(0)^t$; $e_2(t,m) = \frac{1}{me_1(0)} \sum_{k=1}^{\min(M-1,m)} (kt - m + k) e_1(k) e_2(m-k)$.

Proof. For notational convenience, in (5.6) we denote $\alpha_k = \frac{\sum_{j\in S_r} w_j}{w_k}$. Using a Taylor expansion it is shown in Appendix 5.6.1 that

$$F_{X|k_r^*=k,S_r}(x) = F_{Y_k}(x)^{\alpha_k}$$
$$= \begin{cases} \sum_{t=1}^{\infty} e_3(\alpha_k,t) \sum_{m=0}^{t(M-1)} e_2(t,m) F_{Z_k}\left(\frac{x}{\rho}\right)^{Nt-m} & \alpha_k \neq 1 \\ \sum_{m=0}^{M-1} e_1(m) F_{Z_k}\left(\frac{x}{\rho}\right)^{N-m} & \alpha_k = 1. \end{cases} \tag{5.8}$$

To find $F_X(\cdot)$ we compute the expectation over all possible realization of numbers of users providing feedback and over all users to be selected.

$$F_X(x) = E_{k_r^*}[E_{S_r}[F_{X\big|k_r^*=k,S_r}(x)]]. \tag{5.9}$$

The expectation over k_r^* is found by summing over all users and the expectation over S_r is found by considering all possible subsets of users which contain user k in the set of users providing feedback. Combined with (5.4) and $\Pr\{k_r^* = k | k \in S_r\} = \frac{w_k}{\sum_{j\in S_r} w_j}$, we have the desired result in (5.7). $\square$

The conditional density of the SNR on the link to the selected user can now be readily computed and is given by (details in appendix 5.6.1),

$$f_{X|k_r^*=k,S_r}(x)$$
$$= \begin{cases} \alpha_k \sum_{t=1}^{\infty} \frac{e_3(\alpha_k,t)}{t+1} \sum_{m=0}^{(t+1)(M-1)} e_2(t+1,m)(F_{Z_k}(x)^{(t+1)N-m})' & \alpha_k \neq 1 \\ \sum_{m=0}^{M-1} e_1(m)(F_{Z_k}(x)^{N-m})' & \alpha_k = 1 \end{cases} \tag{5.10}$$

An analytical form for the system throughput for the general channel model case is not tractable. Expression for some specific cases is available. In the general case, the system throughput is found by evaluating the integral $\int_0^{\infty} \log(1+x) f_{X|k_r^*=k,S_r}(x)dx$ numerically and then summing up over all possible realization of set of users to feed back as well all possible options of which user to be selected.

5.3.2 Specialization to an OFDMA system with Rayleigh Fading Channel

We have discussed analysis steps and found CDF and PDF of system throughput and SNR as in (5.7) and (5.12), without assuming any specific channel distribution between the BS and the users. Herein, we apply the above frame work to specific system models. At first, an OFDMA system is considered where the channel from the BS to users are under Rayleigh fading.

Specialized Expressions for the Rayleigh fading Channel

In this model, CQI from each user k on resource block r to the base station is $Z_k = |h_{k,r}|^2$, which has exponential distribution.

$$F_{Z_k}\left(\frac{x}{\rho}\right) = 1 - e^{-\frac{x}{\rho c_k}}, \tag{5.11}$$

where c_k is the pathloss and $h_{k,r} \sim CN(0, c_k)$. This CDF is then substituted in (5.7) to obtain the corresponding CDF of the system throughput. The PDF of SNR in (5.12) is

$$f_{X|k_r^*=k,S_r}(x) = \begin{cases} \alpha_k \sum_{t=1}^{\infty} \frac{e_3(t)}{t+1} \sum_{m=0}^{(t+1)(M-1)} e_2(t+1,m) & \\ \quad \times \sum_{s=1}^{(t+1)N-m} \frac{s}{\rho c_k} \binom{(t+1)N-m}{s} (-1)^{s+1} e^{-\frac{sx}{\rho c_k}} & \alpha_k \neq 1 \\ \sum_{m=0}^{M-1} e_1(m) \sum_{s=0}^{N-m} \binom{N-m}{s} (-1)^{s+1} \frac{s}{\rho c_k} e^{-\frac{sx}{\rho c_k}} & \alpha_k = 1 \end{cases} \tag{5.12}$$

The average throughput can be computed in closed form in this specific case. The average thoughput conditioned on the set of users who provide feedback is defined as

$$\bar{R}_{k_r^*=k,S_r} = \int_{x=0}^{\infty} \log(1+x) f_{X|k_r^*=k,S_r}(x)dx. \tag{5.13}$$

From [97, 4.337.2], the integration $\int_0^{\infty} e^{-\beta x} \log(1+x) = -\frac{1}{\beta} e^{\beta} Ei(-\beta)$. Finally, taking expectation over all possible case of the selected user k and all feedback scenarios similar to (5.9), we have

$$\bar{R} = \frac{1}{\ln 2} \sum_{k=1}^{K} \sum_{\pi(S_r|k \in S_r)} \frac{w_k}{\sum_{j \in S_r} w_j} \left(\frac{M}{N}\right)^{|S_r|} \left(1 - \frac{M}{N}\right)^{K-|S_r|}$$
$$\times \begin{cases} \sum_{t=1}^{\infty} \frac{e_3(t)}{t+1} \sum_{m=0}^{(t+1)(M-1)} e_2(t+1, m) \\ \quad \times \sum_{s=1}^{(t+1)N-m} \binom{(t+1)N-m}{s} (-1)^s e^{\frac{s}{\rho c_k}} Ei\left(-\frac{s}{\rho c_k}\right) & \alpha_k \neq 1 \\ \sum_{m=0}^{M-1} e_1(m) \sum_{s=0}^{N-m} \binom{N-m}{s} (-1)^s e^{\frac{s}{\rho c_k}} Ei\left(-\frac{sx}{\rho c_k}\right) & \alpha_k = 1 \end{cases} \tag{5.14}$$

Numerical results

To illustrate the analysis, we examine an OFDMA system with $N = 10$ resource blocks. We consider a group of K_1 users at the middle of the cell and a group of K_2 users at the cell edge with the corresponding weights $w = [0.4, 0.6]$. This configuration is adopted with a view towards comparison with the relay configuration to follow. The BS uses a power of 37dBm to transmit to the user groups located $d_1 = 414$m and $d_2 = 834$m respectively. NLOS channel model E as in [96] is used. Each user feeds back M best channels among its channels on all resource blocks. The system throughput performance is analytically evaluated. One challenge with evaluating the analytical expression is that (5.8) contains an infinite sum which can not be computed exactly. However, we can replace the infinite sum by a finite sum as the Taylor expansion of the formula $(1 + x)^{\alpha_k}$ converges fast when $x < 1$. Numerically, the error is bounded by the term with highest order l_u, $\binom{\alpha}{l_u+1} x^{l_u+1}$, which can be made arbitrarily small by choosing a suitably value of l_u. The analytical results agree very well in all cases studied with the numerical simulations.

Fig. 5.2 shows the relation between the average system throughput and transmit power corresponding to different choices of feedback parameter M. When M increases from 1, the system throughput is enhanced significantly mainly due to the probability that no user provides feedback on a resource block is reduced. However, this improvement is not necessarily monotonic in M as the performance also depends on other factors such as the pathloss c_k, weight w_k for $k = 1, \ldots, K$ and so on. When M is significant, which is about $M > 3$ in this experiment,

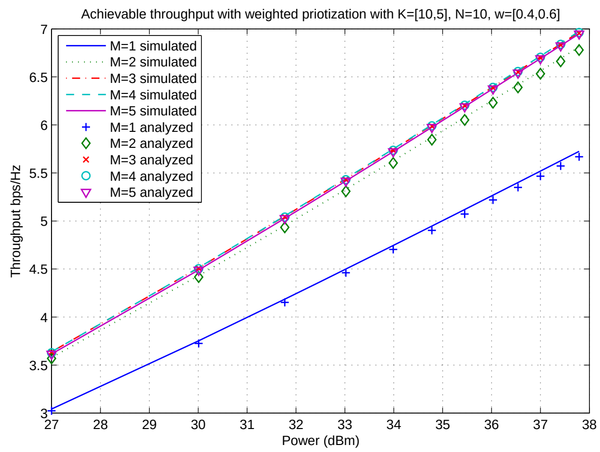


Figure 5.2: Analyzed and simulated performance of a partial feedback OFDMA system with $N = 10$, $K = [10, 5]$.

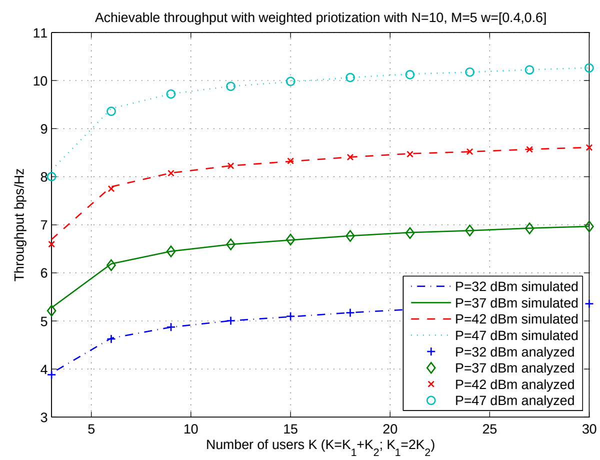


Figure 5.3: Analyzed system throughput of a partial feedback OFDMA system with $N = 10$, $M = 5$, $K_1 = 2K_2$.

the throughput of the system using best M feedback converges to that using full feedback. In Fig. 5.3, the relation between the average system throughput and number of users are shown. In this experiment, the feedback per user is fixed at $M = 5$ while the number of users in the first cluster is set to be twice as large as the second one. When the total number of users increases, the average system throughput increases as the BS can exploit more multiuser diversity. The increment in the average system throughput when adding more users is observed to be more significant with small number of users.

We also analyze system throughput as a function of amount of feedback

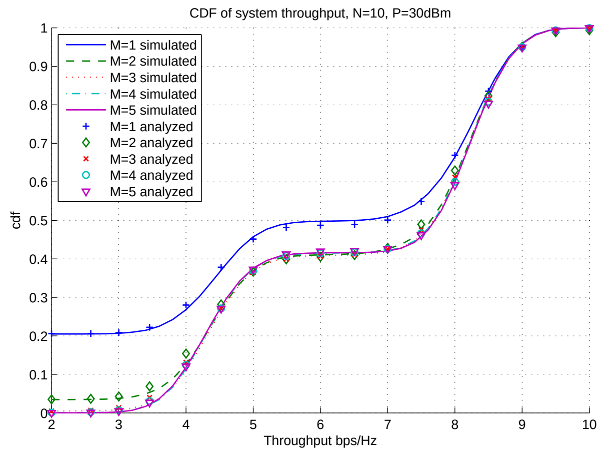


Figure 5.4: CDF of the systems' throughput as a function of number of feedback; OFDMA system with $N = 10$, $K_1 = 10$, $K_2 = 5$, $w = [0.4, 0.6]$.

M. To emphasize the impact of feedback, we assumed the channel is not used when there is no user who provides feedback on that channel. In reality, we can assign that resource block to any user and improve performance. Thus, each curve begins with the CDF equal to the probability that there is no user to provide feed back which equals $(1 - \frac{M}{N})^K$. The result in Fig. 5.4 shows that the amount of feedback beyond a sufficient value of M does not have significant impact on the system throughput. When M is small, the probability of having no feedback degrades the overall performance. When M is large enough to ensure that there is likely at least one user providing feedback, further increase in M does not cause considerable improvement in system performance.

5.3.3 Setting weights for groups of users

Though with the preassigned weights, the allocation for users is controlled, we typically do not know the weights for the groups of users before hand. In reality, the network has resource requirements for users in the form of average throughput and QoS. Suppose one user needs to be allocated two times more resources than another user. Then, the network maps these requirements into suitable priorities for users. From (5.5), the probability of selection of user k is

$$\Pr\{k_r^* = k\} = \sum_{\pi(S_r|k\in S_r)} \frac{w_k}{\sum_{j\in S_r} w_j} \left(\frac{M}{N}\right)^{|S_r|} \left(1-\frac{M}{N}\right)^{K-|S_r|}. \tag{5.15}$$

We know the optimal point when there is full feedback is $w_k(0) = \Pr_{alloc,k}$ where $\Pr_{alloc,k}$ is the needed allocation of a user k. Because the partial feedback changes the probability of selection of users, we find a suitable weight for that user. Let's start our search from $\mathbf{w}(0)$ and expect the algorithm to converge to the optimal value of $\mathbf{w}$. For this purpose, we construct a function $\Phi(\mathbf{w})$ which has derivative $[\phi_1, \ldots, \phi_K]^T$ with $\phi_k(\mathbf{w}) = \Pr\{k_r^* = k\} - \Pr_{alloc,k} = 0$ for all $k = 1, \ldots, K$. This condition corresponds to the global optimum of $\Phi(\mathbf{w})$.

If we can show the 2nd derivative of $\Phi(\mathbf{w})$ or equivalently the derivative of $\boldsymbol{\phi}(\mathbf{w})$ is positive definite, then by the convexity of $\Phi(\mathbf{w})$ we can conclude that the optimum value of $\Phi(\mathbf{w})$ is at $\mathbf{w}^*$ and $\mathbf{w}^*$ is unique. The details are relegated to the appendix 5.6.2. A zero finding algorithm [98] to find weight $\mathbf{w}$ is as follows

- Initialize $\mathbf{w}(0) = [\Pr_{alloc,1}, \ldots, \Pr_{alloc,K}]$, which is the weights for the full feedback case.
- Solve $\nabla\boldsymbol{\phi}(\mathbf{w}(t))\delta\mathbf{w}(t) = -\boldsymbol{\phi}(\mathbf{w}(t))$.
- Update $\mathbf{w}(t+1) = \mathbf{w}(t) + \delta\mathbf{w}(t)$. Normalize $\mathbf{w}$ so that $\|\mathbf{w}\|_1 = 1$ without changing $\boldsymbol{\phi}(\mathbf{w})$.
- End $\|\boldsymbol{\phi}(\mathbf{w})\|_2 < \epsilon$.

In this experiment, by applying this zero finding algorithm for a partial feedback OFDMA system with $N = 10$, $M = 5$, $K = 15$, $\epsilon = 10^{-10}$, and the targeted probability of selection for users in the two clusters proportional to $\frac{4}{6}$, we find weight for each of the 10 users in Cluster 1 is 0.0365 and that in cluster 2 is 0.127. The algorithm converges after 4 iterations as tabulated in Table 5.1.

Table 5.1: Convergence of the weight finding algorithm. The norm as a function of number of iterations.

Iteration	Norm $\|\boldsymbol{\phi}(\mathbf{w})\|_2$
0	9.1539269e-02
1	8.8913174e-03
2	7.7456533e-05
3	5.9135413e-09
4	2.7755576e-17

5.4 OFDMA systems with both groups of macro users and relays

In this section, we consider to include relays to the system where the relays are deployed to improve network coverage at the cell boundaries. Herein, DF relays are assumed and the transmission has two phases. In the first interval, the BS transmits to the selected relay. The relay decodes the received message and forwards it to the served user. Because the system has multiple resource blocks, there are options for the transmission on the two hops.

- The transmissions on BS-relay and relay-user links are allowed to be on two different resource blocks. This option can potentially exploit frequency diversity to obtain good communication to the user. However, this option requires more feedback to the BS as the BS needs to know frequencies matching on all the relays in order to avoid frequency conflict on the $2nd$ hop.
- The transmissions on BS-relay and relay-user links are on the same resource block. This way, the BS can always make sure all the communications on the $2nd$ hop are on disjoint resource block. As we have many users in each relay groups, the multiuser diversity ensures the link on the $2nd$ hop has good quality.

In this work, we the simpler option is investigated: the communications between the BS and the relay as well as between the relay to the corresponding user are assumed to be on the same resource block. The other options will be explored in our future work.

5.4.1 System 's feedback requirement

One important factor related to the introduction of the relays is each relay represents a group of users it serves. Hence, it often requires more resource and should be assigned a larger weight. Naturally, an arisen question is whether the partial feedback scheme is able to support the resource requirement of the relays. Let's look at the time share portion of a relay l which serves K_l users.

From (5.5), the probability that the relay is selected is

$$\begin{aligned} \Pr\{k_r^* = k\} &= \sum_{\pi(S_r|k\in S_r)} \frac{w_k}{\sum_{j\in S_r} w_j} \left(\frac{M}{N}\right)^{|S_r|} \left(1-\frac{M}{N}\right)^{K-|S_r|} \\ &\leq \sum_{\pi(S_r|k\in S_r)} \left(\frac{M}{N}\right)^{|S_r|} \left(1-\frac{M}{N}\right)^{K-|S_r|} = \frac{M}{N}, \end{aligned} \tag{5.16}$$

where the 2nd step follows from $\frac{w_k}{\sum_{j\in S_r} w_j} \leq 1$. Equivalently, we have the condition on partial feedback $\frac{M}{N} \geq \Pr\{k_r^* = k\} \quad \forall k$ or

$$\frac{M}{N} \geq \max_k \Pr\{k_r^* = k\}. \tag{5.17}$$

5.4.2 System throughput

Assuming the mentioned condition on partial feedback is satisfied. We want to investigate how to select a value M so that the system throughput loss is minimal. Herein, we want to know if M should be significantly or just slightly larger than the minimum requirement in (5.17). From equation (5.14), we investigate the average rate of a user k as a function of $\frac{M}{N}$.

For an OFDMA system which contains both groups of macro users and relays, the distribution $F_{Z_l}(x)$ in (5.7) is now the CDF of the overall SNR of the link through the relay. Users served by a relay view the relay as their base station. We denote the SNR on the link from the BS to the relay by $W_{1,l}$ and the SNR from the relay to an arbitrary user k and to the selected user by $W_{2,l,k}$, $W_{2,l}$ respectively. For notational simplicity, we do not show the resource block r on the variables as the statistics on every resource block are assumed identical. The overall CDF of the SNR if the BS selects relay l and the relay selects users k on that resource block, is determined by the minimum SNR of the two hops and is given by [53, 2.1.2]

$$F_{Z_l}(x) = 1 - (1 - F_{W_{1,l}}(x))(1 - F_{W_{2,l}}(x)). \tag{5.18}$$

When a relay receives feedback from users it serves, it compares the CQI of the link to the selected user (2nd hop) with the CQI on the link on the 1st hop and feeds back the minimum value to the BS.

The link BS-relays are under Rayleigh fading

When the links from BS to relays and from relays to users are under Rayleigh fading and have the same channel coherence time, the variables $W_{1,l}$ and $W_{2,l,k}$ have exponential distribution. The CDF of $W_{1,l}$ is $F_{W_{1,l}}(x) = 1 - e^{-\frac{x}{\rho_{1,l}c_{1,l}}}$. Because the users served by a relay are far apart from the BS, we can assume these user to reuse some frequency resource to provide full feedback to the relay. This assumption avoids the complication come from the fact that the best M CQI on the 2nd hop do not correspond to the best CQI of the first hop. Herein, we derive throughput of a relay when it serves a single user. Similar result can be obtained when the relay serves more users.

The system throughput is found by plugging the distribution of $F_{Z_l}(x)$ into the throughput distribution equation in (5.12) and taking the integration in (5.13). However, we do not present the result mainly due to the computational complexity of the calculation. Instead, we are more interested in how the partial feedback affects the overall performance. Herein, a simple case where a relay l serves only one user k is considered. In order for the relay to choose the best M sub-carriers, the user provides full feedback to the relay. As the user k is assumed isolated from the BS, it can reuse uplink resource to provide full feedback without affecting system performance. The distribution of the overall channel CQI is (5.18)

$$F_{Z_l}(x) = 1 - (1 - F_{W_{1,l}}(x))(1 - F_{W_{2,l}}(x)) = 1 - e^{-\frac{x}{\rho_l}}, \tag{5.19}$$

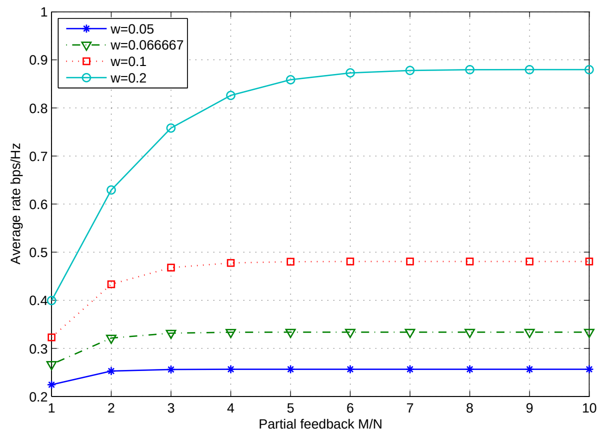


Figure 5.5: Throughput of a user k as a function of partial feedback parameter.

with $\rho_l = \frac{\rho_{1,l}\rho_{2,l,k}}{\rho_{1,l}+\rho_{2,l,k}}$.

To show how the partial feedback causes negative effect to system throughput, a system where every users are assigned the same weight is considered. The throughput of the communication to user k through relay l is shown. The equivalent SNR is set $\rho_l = 10$ dB. For the chosen weights, the minimum feedback parameters M are between 1 and 2. However, even with M larger than this minimum requirement (5.17), significant loss in throughput is observed as shown in Fig. 5.5. This experiment suggests us to select a large enough M not only to satisfy the minimum feedback requirement shown in (5.17) but to ensure negligible loss in user's throughput. The minimum M can be set $M \geq \max\left(1, N(1-\epsilon^{\frac{1}{K}}\right)$ as in [99] to ensure feedback availability ϵ on each sub-carrier.

In Fig. 5.6, we investigate system performance when there are both macro users and relays. The configuration is the same as in Fig. 5.4 but we replace group 2 with a relay at the same location. The distance from the BS-relay and relay-user are set the same. The link BS-relay is considered Rayleigh fading with NLOS connection as in Model E in [100]. In the system without a relay, the overall performance is lower. In the figure, the throughput from 3 to 5bps/Hz corresponds to the allocation for the users at the cell edge. When a relay is added, system throughput is enhanced significantly even when the communication through the relay is conducted in 2 phases which results in a factor of $\frac{1}{2}$ in the rate equation.

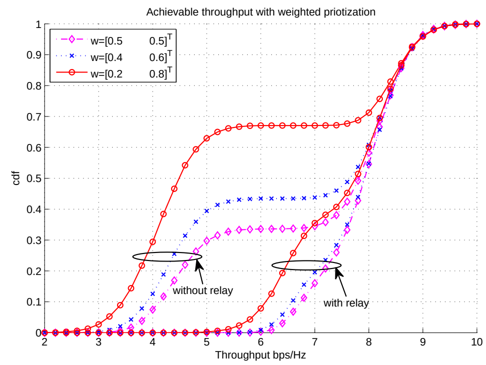


Figure 5.6: Performance tradeoff due to the biased treatment with users; OFDMA system with $N = 10$, $K = [10, 5]$, $P = 37$dBm, the distance BS-macro user is $d_1 = 414$m and the distance BS-user group 2 is $d_2 = 834$m which are aided by a relay with power 30dBm located 815m from the BS, full feedback is provided.

Moreover, the involvement of the relay can possibly help the BS to save transmit power which is investigated further in section 5.4.3. The analytical expressions also agree very well with the numerical results.

5.4.3 Resource allocation

For channel with fast fading, resource is allocated to a relay in a similar way as to a macro user. In reality, the links between the BS and relays are generally predesigned and fixed so the channels between them maybe slowly varying. A natural question that arises is how these slowly varying channels, i.e. different channel coherence times, affect the performance of the weighted CDF based scheduling.

We study this experimentally to develop some intuition and then suggest some potential solutions to mitigate the shortcomings. The distribution of the minimum SNRs on the two hops shown in (5.18) is without consideration to coherence time. To magnify the effect, consider the case where the first hop is nearly constant over a reasonable time period. Over that period, due to the unchanged SNR g_l on the 1st hop of relay l, the CDF of the overall SNR on this link is

$$F_{Z_l|W_{1,l}=g_l}(x) = \begin{cases} F_{W_{2,l}}(x) & x<g_l \\ 1 & x\geq g_l \end{cases}. \tag{5.20}$$

Note that $W_{2,l}$ or equivalently W_{2,l,k^*} is a random variable representing the SNR on the link from relay l to the selected user k^*. The CDF of $W_{2,l}$ depends on certain resource allocation and feedback scheme between the users in cluster l to its relay. To avoid unnecessary complication and focus on the dynamic of the two hops, we can assume full feedback on the 2-nd hop. There are several options to calculate the CDF metric as follows

- Using conditional CDF: When we directly apply the CDF in (5.20) to our CDF based scheduler, the user with the poorest channel on the 1st hop is served frequently while the user with better channel quality is not. This is can be explained by observing that the CDF of the user with the poorest channel on the 1st hop often gets value 1 and so it has the highest CDF comparison metric and is served more frequently.

- Long term fairness: Another option is to apply directly the CDF in (5.18) incorporating the large coherence time of the 1st hop. This approach favors users in the clusters with better 1st links because the computed CDF of the others users never exceed the CDF evaluated at g_l which has a small value.

- Short term fairness: We use an extra interpolation step to the conditional CDF as will be discussed in detail later.

This suggests a potential problem in dealing with users with differing channel coherence time. The CDF based scheduler is fair in the long term. However, fairness in the short term is only assured if the channels are i.i.d over time. With relays this is not the case and so we explore options to provide short term fairness at the expense of system throughput.

Short term fairness

This scheme is based on some insight that have been employed for CDF based scheduling with quantized CQI. We substantiate them by showing their fairness guarantees. Because they are CDF based, one can expect them to share some of the diversity benefits but no optimality in the fairness-diversity tradeoff can be claimed. The simulations results are promising and indicate an area for future exploration.

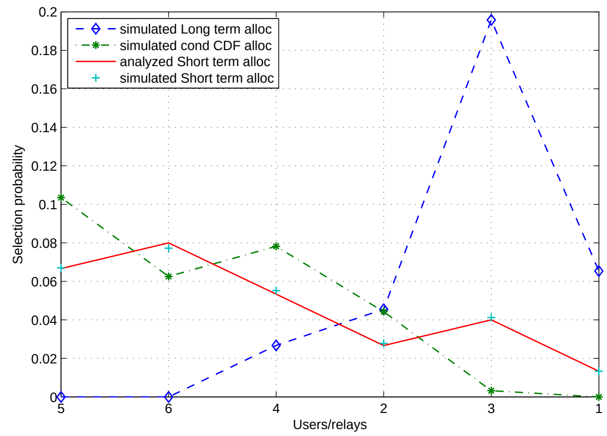


Figure 5.7: Selection probability of users in an OFDMA system with $L = 5$ relays, Cluster 1 contains macro users, number of users in clusters are $K = [2\ 2\ 2\ 3\ 3\ 3]$, weight $w = [0.05\ 0.1\ 0.15\ 0.2\ 0.2\ 0.3]$, BS-macro user group distance are 414m, and BS-cell edge users distance are 834m, BS-relays distance are 815m, Log-normal fading 8dB.

In this scheduling scheme, we replace the jump in the CDF by a continuum of values containing all transition values of $F_{Z_l|W_{1,l}=g_l}(x)$ from $F_{Z_l|W_{1,l}=g_l}(g_l)$ to 1. When $W_{2,l} \geq g_l$,, the values for $F_{Z_l|W_{1,l}=g_l}(x)$ instead of being set at 1 are generated using an independent random variable which is uniformly distributed between $F_{Z_l|W_{1,l}=g_l}(x)$ and 1. As $F_{W_{2,l}}(x)$ is uniformly distributed, we can use this CDF for the constructed transition in CDF of $F_{Z_l|W_{1,l}=g_l}(x)$ at $x = g_l$. As we use the relay to represent all users it serves, we consider it as a user l with weight w_l. Similar to (5.5), the probability relay l is selected is

$$\Pr\{k^* = l | x_l = x\} = \prod_{k \neq l, k=1,\ldots,K} \Pr\left\{u_k^{\frac{1}{w_k}} < u^{\frac{1}{w_l}}\right\} = u^{\sum_{k=1}^{L} \frac{w_k}{w_l} - 1}, \tag{5.21}$$

where u_k is a random variable uniformly distributed in $[0, 1]$ which is the result of a mapping from channel CQI using the CDF function $F_{Z_k}(x_k)$. Taking expectation over value of x which makes the corresponding u vary from 0 to 1, we have the probability that relay l is selected

$$\Pr\{k^* = l\} = \int_0^1 u^{\sum_{k=1}^{L} \frac{w_k}{w_l} - 1} du = w_l. \tag{5.22}$$

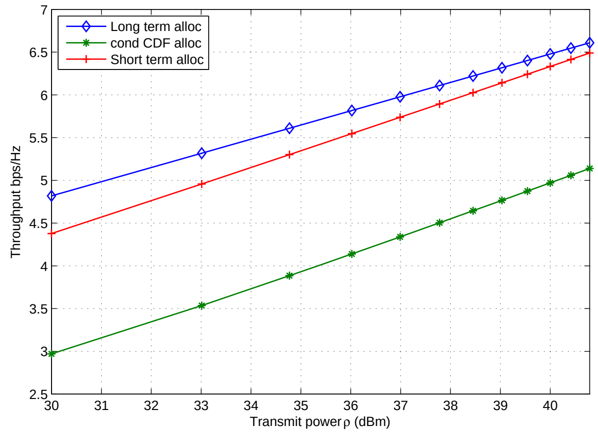


Figure 5.8: Tradeoff in system's throughput when adjusting weight of the user in outage condition. OFDMA system with $L = 5$ relays, group 1 contains macro users, number of users in clusters are $K = [5\ 3\ 3\ 3\ 3\ 3]$, weight $w = [1\ 2\ 3\ 4\ 5\ 6]$, BS-macro user group distance are 414m, and BS-user in relay groups distance are 834m, BS-relays distance are 815m, Log-normal fading 8dB.

This way, we are still able to proportionally allocate resource to all the users as shown in Fig. 5.7, but with some reduction in overall system performance as shown in Fig. 5.8.

To illustrate the resource starvation resulting from the previous schemes due to the differing channel coherence times in the two links and to illustrate the benefits from the suggested scheme, a system with 5 relays and macro users is considered. There are 5 macro users who are considered identical and are grouped into Cluster 1. The purpose of this grouping of macro users is to simplify the presentation result as it allows us to focus more on the impact of relays. Each relay can be considered a single equipment or a cluster interchangeably. Number of users and the weights for these clusters are $K = [5, 3, 3, 3, 3, 3]$ and $[1, 2, 3, 4, 5, 6]$ correspondingly. The weights are then normalized so the sum of weights of all the users is equal to 1. Note we can simply consider all macro users as users served by relays with infinite SNR on the $1st$ hop. The SNRs on the $1st$ hop is generated according to a log-normal distribution with variance 8dB which have value $[\infty, -44.0, -66.3, -62.0, -68.7, -91.4, -77.6]$ dB. The match between the selection probability of a user in the groups as predicted by the analysis (5.22) and the simulated one is shown in Fig. 5.7. The agreement conforms the analysis.

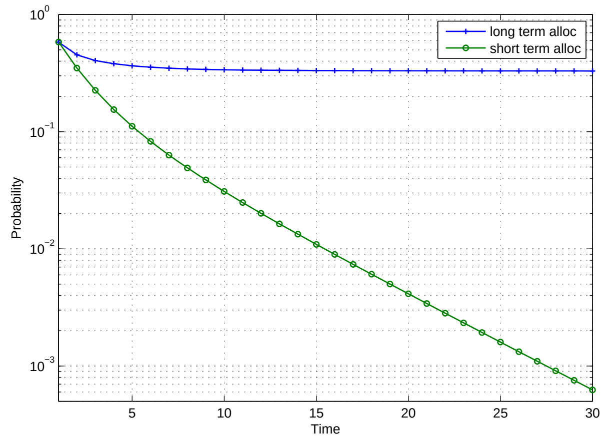


Figure 5.9: Average fraction of users who are not allocated resource after t intervals using Long term allocation. OFDMA system with $L = 5$ relay, Cluster 1 contains macro users, number of users in clusters are $K = [5\ 3\ 3\ 3\ 3\ 3]$, weight $w = [1\ 2\ 3\ 4\ 5\ 6]$, BS-macro user group distance are 414m, and BS-user in relay groups distance are 834m, BS-relays distance are 815m, Log-normal fading 8dB.

Another effect of the slow fading BS-relay link is a fraction of users are not allocated for one or even many scheduling periods. This worst case scenario happens when insufficient frequency diversity, slow and deep fading all occur at once. Typically, if a user is not allocated after T_{max} intervals, it is considered be in resource starvation condition. We call the user to be in deadline condition. Assuming there is no frequency diversity, the average probability that users in the system are in deadline condition is shown as a function of T_{max} in Fig. 5.9. On average, the conditional CDF approach and the use of actual CDF result in severe starvation probability. Moreover, using these methods we can not meet the targeted allocation for users in a short term manner. Meanwhile, the modified CDF developed in the paper has very good capability of preventing resource starvation as shown in Fig. 5.9 and ensures short term fairness. It also offers good system performance as can be seen in Fig. 5.8, indicating that the tradeoff is made in a reasonable way.

Power saving scheme

The method developed for ensuring short term fairness has some flexibility. The auxiliary uniform random variable generated between g_l and 1 is generated

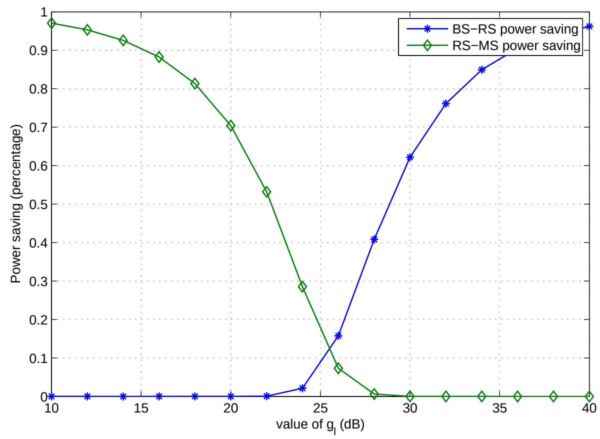


Figure 5.10: Percentage of power saving on the two links of a relay l, with weight $w_l = 0.1$, distance BS-relay is 815m, relay-user is 50m.

independently and can potentially be coupled with other metrics as discussed next. As the equivalent SNR on the link BS-relay-user is defined by the minimum SNR of the two links, using extra power on the better link does not help in increasing throughput of the communication. Thus, we propose to use the auxiliary random variable to help reduce transmit power on the better link so that the two links are still able to support the required transmission rate. This power back-off mechanism helps to save the power used for the link. Instead of using a totally random auxiliary variable as in section 5.4.3, we use the CDF of the SNR on the second hop to create the transition from $F_{Z_l}(g_l)$ to 1, i.e. users are selected based on the weighted CDF of $F_{W_{2,l}}(x)$. This results in improved power savings as explained next. Note that given the SNR on the BS-relay link g_l is constant, the overall SNR of the link is $\min(g_l, x)$ where x is the instantaneous SNR of the user selected by relay l. The transmission conditions has two possibilities as described below.

- The $x < g_l$, then the 1st link (BS-relay) can back off its transmit power

$\hat{\rho}_{1,l} = \frac{\rho_{1,l} x}{g_l}$ which results in a power saving of

$$P_1 = \int_0^{g_l} \left(\rho_{1,l} - \hat{\rho}_{1,l}\right) F_{W_{2,l}}(x)^{\frac{1}{w_l}-1} f_{W_{2,l}}(x) dx = w_l \rho_{1,l} \left(1 - e^{-\frac{g_l}{\rho_{2,l} c_{2,l}}}\right)^{K_l} - \frac{w_l \rho_{1,l} \rho_{2,l} c_{2,l}}{g_l} \sum_{m=1}^{\infty} \binom{K_l/w_l}{m} \frac{(-1)^{m+1}}{m} \left(1 - e^{-\frac{m g_l}{\rho_{2,l} c_{2,l}}} \left(1 + \frac{m g_l}{\rho_{2,l} c_{2,l}}\right)\right), \tag{5.23}$$

where the result is derived using Taylor expansion for $F_{W_{2,l}}(x)^{\frac{1}{w_l}-1}$, similar to the derivation in appendix 5.6.1. Then, the integration of the exponential functions is conducted in a straightforward manner.

- The $x \geq g_l$, the 2nd link (relay-user) can back off its transmit power $\hat{\rho}_{2,l} = \frac{g_l \rho_{2,l}}{x}$. Though the overall SNR can not exceed g_l, from a power saving point of view, it is better if the channel is used when the SNR on the second hop is large as facilitated by this scheme. The power saving is

$$P_2 = \int_{g_l}^{\infty} \left(\rho_{2,l} - \hat{\rho}_{2,l}\right) F_{W_{2,l}}(x)^{\frac{1}{w_l}-1} f_{W_{2,l}}(x) dx = w_l \rho_{2,l} \left(1 - \left(1 - e^{-\frac{g_l}{\rho_{2,l} c_{2,l}}}\right)^{K_l}\right) - \frac{w_l g_l}{c_{2,l}} \sum_{m=1}^{\infty} \binom{K_l/w_l}{m} m (-1)^m Ei\left(-\frac{m g_l}{\rho_{2,l} c_{2,l}}\right), \tag{5.24}$$

where the exponential integral in the last equation is obtained by applying [97, 8.211.1].

In Fig. 5.10, we show the percentage of the power saved on two hops of a relay l with weight w_l. The distances BS-relay, relay-user are set 815m and 50m. When the obtained SNR on the first hop is small, the relay backs off and save around 90% of its power. On the other hand, when the obtained SNR of the slow fading hop increases, the BS-relay link can save on its transmit power. When g_l is large, which means the link BS-relay is in good condition, the relay almost always uses up its power while the BS saves a large portion of its power.

5.5 Conclusion

In this paper, we provide a thorough analysis of a partial feedback OFDMA system with either only macro users or with both macro users and users served by relays. Users in the system are prioritized using weighted CDF based scheduling. We provide an analytical expression for both the average system throughput and the distribution of system throughput. When relays are added to provide services for cell edge users, the difference in time scale on the two hops of each relay is shown to cause resource starvation for users. A modified version of the weighted CDF based scheduling is developed to guarantee short term fairness. Due to the imbalance on the SNR of the two hops of each relay, we also propose a scheme to back off transmit power on either of the hops to save the overall power used on the links of the relay.

The text of this chapter, in full, is a reprint of the paper [101], Anh H. Nguyen, Yichao Huang and Bhaskar D. Rao Weighted CDF based scheduling for OFDMA relay downlink with partial feedback, that is being prepared for publication on IEEE Transaction on Wireless Communications. The dissertation author is the primary researcher and author, and the co-authors listed in this publication collaborated and supervised the research constitutes this chapter.

5.6 Appendices

5.6.1 CDF and PDF of system's throughput

Herein, we consider the nontrivial proof of (5.8), when $\alpha_l \neq 1$. The case $\alpha_l = 1$ is straight forward and no further expansion is needed.

$$
\begin{aligned}
F_{X|k_r^*=k,S_r}(x) &= [1 + (F_{Y_k}(x) - 1)]^{\alpha_k} && (a)\\
&= 1 + \sum_{i=1}^{\infty} \binom{\alpha_k}{i} \left((-1)^i + \sum_{t=1}^{i} \binom{i}{t} (-1)^{i-t} F_{Y_k}(x)^t \right) && (b)\\
&= \sum_{i=1}^{\infty} \binom{\alpha_k}{i} \sum_{t=1}^{i} \binom{i}{t} (-1)^{i-t} F_{Y_k}(x)^t && (c)\\
&= \sum_{t=1}^{\infty} e_3(\alpha_l, t) \left(\sum_{m=0}^{M-1} e_1(m) F_{Z_k}(\frac{x}{\rho})^{N-m} \right)^t && (d)\\
&= \sum_{t=1}^{\infty} e_3(\alpha_l, t) \sum_{m=0}^{t(M-1)} e_2(t,m) F_{Z_k} \left(\frac{x}{\rho} \right)^{Nt-m} && (e)
\end{aligned} \tag{5.25}
$$

where we have (c) as $1 + \sum_{i=1}^{\infty} \binom{\alpha_l}{i} (-1)^i = (1-1)^{\alpha_l} = 0$. We have (d) by replacing $\sum_{i=1}^{\infty} \binom{\alpha_l}{i} \sum_{t=1}^{i} \binom{i}{t} (-1)^{i-t} F_{Y_k}(x)^t = \sum_{t=1}^{\infty} \sum_{i=t}^{\infty} \binom{\alpha_l}{i} \binom{i}{t} (-1)^{i-t} F_{Y_k}(x)^t$ and set $e_3(\alpha_l, t) = \sum_{i=t}^{\infty} \binom{\alpha_l}{i} \binom{i}{t} (-1)^{i-t}$. Finally, we apply [97, 0.314] to get (e) with $e_2(0) = e_1(0)^t$ and $e_2(m) = \frac{1}{me_1(0)} \sum_{k=1}^{\min(M-1,m)} (kt - m + k) e_1(k) e_2(m-k)$. The PDF of the SNR on the link to the selected user is obtained by taking the derivative of the CDF.

5.6.2 The uniqueness of the found weight w

We have assumed that a suitable function $\Phi(\mathbf{w})$, which has derivative $\boldsymbol{\phi}(\mathbf{w})$, has been constructed. This can be done by integrating $\boldsymbol{\phi}(\mathbf{w})$ but an explicit form is not needed. If $\boldsymbol{\phi}'(\mathbf{w})$ is positive definite, then $\Phi(\mathbf{w})$ is convex and has a unique minimizer corresponded to the zero of $\boldsymbol{\phi}(\mathbf{w})$. As a consequence, the found zero of $\boldsymbol{\phi}(\mathbf{w})$ is unique.

From (5.5), the probability of the selection of user k is summed over all possible set of feedback $\pi(S_r)$. As sum of positive definite matrices is positive definite, we need to show the derivative of every component $\left[\frac{w_1}{\sum_{j=1}^{K} w_j}, \ldots, \frac{w_K}{\sum_{j=1}^{K} w_j} \right]^T$

is positive definite

$$\Pr\{k_r^* = k\} = \sum_{\pi(S_r|k\in S_r)} \frac{w_k}{\sum_{j\in S_r} w_j} \left(\frac{M}{N}\right)^{|S_r|} \left(1-\frac{M}{N}\right)^{K-|S_r|}$$
$$= \sum_{\pi(S_r|k\in S_r)} \frac{w_k}{\sum_{j\in S_r} w_j} \theta_{S_r}. \tag{5.26}$$

We want to show the positive definiteness of the $K \times K$ matrix by showing the positive semi-definite of all sub-matrices. For each S_r, the derivative is a $K \times K$ size matrix which have the diagonal elements (k, k) equal $\frac{1}{(\sum_{j\in S_r} w_j)^2} \sum_{j=1;j\neq k}^{K} w_j$, the off-diagonal elements equal $-\frac{1}{(\sum_{j\in S_r} w_j)^2} w_k$ if $k, l \in S_r$ and 0 otherwise. Let $L = |S_r|$, and we consider the sub-matrix A of size $L \times L$ where the elements are from the nonzero elements of the derivative matrix which are at position k, l with $k, l \in S_r$. We ignore the multiplication factor $\frac{1}{(\sum_{j\in S_r} w_j)^2}$ as it does not change the positive definiteness of the matrix. Let $a = \sum_{j\in S_r} w_j$. We find the determinant of any sub matrix of A by using the following step

$$\begin{vmatrix} a-w_1 & -w_2 & -w_3 & -w_4 \\ -w_1 & a-w_2 & -w_3 & -w_4 \\ -w_1 & -w_2 & a-w_3 & -w_4 \\ -w_1 & -w_2 & -w_3 & a-w_4 \end{vmatrix} = \begin{vmatrix} a-w_1 & -w_2 & -w_3 & -w_4 \\ -a & a & 0 & 0 \\ -a & 0 & a & 0 \\ -a & 0 & 0 & a \end{vmatrix} = \begin{vmatrix} \hat{a} & 0 & 0 & 0 \\ -a & a & 0 & 0 \\ -a & 0 & a & 0 \\ -a & 0 & 0 & a \end{vmatrix}. \tag{5.27}$$

where in the $1st$ step, we subtract the $1st$ row from all rows except for the $1st$. Then, we add the $\frac{w_l}{a}$ value of row $l = 2, 3, 4$ to the 1st row, and we have $\hat{a} = a - w_1 - w_2 - w_3 - w_4$. Replacing a with $a = \sum_{j\in S_r} w_j$, we have $\hat{a} = \sum_{j\in S_r; j\neq 1,2,3,4} w_j \geq 0$. As the resulting lower triangular matrix has all diagonal element greater than or equal to zero, it is a semi-positive determinant. All sub-matrix has semi-positive determinant. The largest size sub-matrix A has zero determinant as for this matrix $\hat{a} = 0$. Using Sylvester's criterion, matrix A is positive semi-definite.

As linear combination of positive definite matrices is positive definite, $2nd$ derivative of $\Phi(\mathbf{w})$ is positive definite. Thus, $\Phi(\mathbf{w})$ is convex and the found root $\mathbf{w}$ is the unique root.

Table 5.2: A single cell multiuser system: the parameters used for the experiments.

BS TX maximum power	47 dBm
Relay TX maximum power	30 dBm [102]
BS-relay channel model	IEEE 802.16j, Type H [96]
BS-user channel model	IEEE 802.16j, Type E [96]
Relay-user channel model	IEEE 802.16j, Type E [96]
Cell radius	876 m
Carrier frequency	2 GHz
Noise power	-144 dBW
Mobile height	1 m
Relay height	15 m
BS height	30 m
Propagation environment	Urban

5.6.3 Simulation Model

To illustrate our results, we use a channel model similar to that in [96] and [100]. The parameters are summarized in Table 5.2 for easy reference.

Chapter 6

Novel Partial Feedback Schemes and their evaluation in an OFDMA system with CDF based scheduling

In an Orthogonal Frequency Division Multiple Access (OFDMA) system, cumulative distribution function (CDF) based scheduling can be applied to prioritize users. In such a wideband multiuser system, it is essential to reduce the typically very large feedback requirement. Herein, we consider a system in which several adjacent highly correlated subcarriers are grouped into a resource block. Then, the formed resource blocks can be further combined to form subbands, each consists a certain number of resource blocks. To reduce feedback, three feedback schemes, which are the opportunistic, the best M and the hybrid schemes, are considered and compared. We provide analytical results which are used to optimally set feedback parameters to best exploit system resource while still maitain fairness among users. Beside, the convergence behaviors of the total feedback requirement of the schemes are also provided. Then, the simulation is conducted to verify the tradeoff between feedback reduction and system performance. The experiments also show the optimal hybrid and opportunistic schemes are identical and result in the least feedback requirement.

6.1 Introduction

In wireless communications, OFDMA is a widely used technique in wideband systems as it can efficiently combat against frequency selective fading. In order to exploit multiuser diversity as well as conduct scheduling techniques, the base station (BS) needs knowledge of the channels to the users, which is obtained through feedback [23]. Due to scheduling policy, the BS assign each resource block for a user. The user to be selected for resource, has highest channel quality in opportunistic scheduling or already has its periodically predetermined allocation in round robin scheduling. These two methods, one favors the user with the best channel condition while the other can not efficiently exploit multiuser diversity which causes loss in system performance. Thus, a scheduling policy that exploits multiuser diversity while maintain fairness among users, is desirable.

CDF based scheduling first proposed in [21] is shown to be able to control precisely allocation for each of the users. Unlike proportional fair scheduling in which system performance is hard to analyze due to the mutual dependance in the selection of users, in CDF based scheduling throughput of the user can be decoupled and hence analytically tractable. The performance of CDF based scheduling is further investigated under best M feedback in [49]. Both of the two works in [21, 49] assume explicit feedback information at the BS.

Though perfect feedback is desirable, such feedback requirement is impractical and requires a huge amount of system resource to obtain. In real system, only imperfect or partial feedback [83, 103] can be obtained at the BS. To reduce feedback a threshold based feedback is proposed in [71] and a contention based feedback scheme [104]. In [85], several feedback schemes are shown to have influence on both feedback requirement and system performance. Moreover, the selection of feedback parameters such as the threshold in the sequential scheme and the feedback rate in contention scheme are very important. In [72], system performance is analyzed under the assumptions of partial feedback and erroneous feedback.

In [105], performance of different feedback schemes are investigated. Under the constraint on the availability of feedback at the BS, parameters of the feedback schemes are set up in order to best reduce feedback requirements [105, 106]. The system throughput with quantized feedback under a M-QAM technique is evaluated in [107]. Beside, performance of OFDMA system using best M feedback scheme is carefully analyzed in [84]. Based on channel characteristic, grouping multiple resource blocks into a subband can further reduce feedback requirement [68] with the tradeoff in a slight reduction in system throughput. Meanwhile the performance analysis of the CDF-based scheduling in a generic multicell is first provided in [48]. Motivated from the context of these works, we consider an OFDMA downlink system wherein we are able to control fairness among users, and best exploit multiuser diversity with the help CDF based scheduling. The needed channel information is obtained from the users through feedback. We group resource blocks to reduce feedback, and examine performance of different feedback schemes which are opportunistic, best M and hybrid scheme. Then, the optimal parameters for the schemes are derived which help compare and pick up the best scheme.

Finally, the convergence behavior of the feedback requirement corresponded to the feedback schemes when number of users go to infinity is observed. Interestingly, the derived analytical formula shows when the number of users goes to infinity, the feedback requirement of the opportunistic and hybrid schemes converge to a certain value.

6.2 System model

We consider an OFDMA downlink system where the bandwidth is divided into N_c sub-carriers. The BS with a single antenna communicates with K users, each has a single antenna. In this work, we consider a delay spread wide band channels with L taps. Then, we obtain N equivalent channels in frequency domain

$$h_{k,r} = \sum_{l=0}^{L-1} \sigma_l F_{k,l} \exp\left(-\frac{j2\pi lr}{N_c}\right), \tag{6.1}$$

where $h_{k,r}$ is the channel to user k on resource block r, L is the number of taps, and σ_l for $l = 0, \ldots, L-1$ represents the channel power delay profile and is normalized $\sum_{l=0}^{L-1} \sigma_l^2 = 1$. The variable $F_{k,l}$ denotes the discrete time channel impulse response, which is a complex Gaussian distributed random process with zero mean and unit variance $CN(0,1)$ and is i.i.d. across k and l.

As the channel on the resource blocks are typically correlated, R_c adjacent sub-carriers are grouped into a resource blocks. Then, η adjacent resource blocks can be grouped into a subband to reduce feedback [68]. The value of η depends on channel characteristic as if the channels are highly correlated, η can be chosen large. Beside, when η gets larger, the error in the fed back channel information becomes larger which degrade system performances. This way, the system has $N = \frac{N_c}{R_c}$ resource blocks and $\frac{N}{\eta}$ subbands.

The average channel throughput on a subband r is defined [68] by

$$S_{k,r} = \sum_{n=(r-1)\eta R_c+1}^{r\eta R_c} \log_2\left(1 + \rho |h_{k,r}|^2\right), \tag{6.2}$$

where ρ is the transmit SNR, assuming the pathloss is already compensated by the BS. From now on, we consider $S_{k,r}$ as the representation of channel to user k on subband r in the analysis because of the one-one mapping between it and the corresponded average SNR.

On each subband r, the BS select one among the fed back users based on the CDF based scheduling [21, 68]. We denote Z_k a random variable which represents the achievable throughput to user k. Assume the BS has the fed back achievable throughput x_k on the link to user k, it obtains the corresponding CDF of that throughput $F_{Z_k}(x_k)$. To support priorities, each user k is assigned a weight w_k proportioned to its priority with $\sum_{k=1}^{K} w_k = 1$. Using CDF based scheduling, the BS will assign this subband r to the user with highest value of weighted CDF.

$$k_r^* = \arg\max_k F_{Z_k}(x_k)^{\frac{1}{w_k}}. \tag{6.3}$$

To further reduce feedback, three feedback schemes which are the opportunistic, best M and the hybrid scheme are considered in the followed sub-sections.

6.2.1 Partial opportunistic feedback

As the analysis on channels on the resource blocks in an OFDMA system are similar, we analyze feedback behavior of user on an arbitrary subband r. In CDF based scheduling [21], a user k is selected if it has the highest weighted CDF $F_{Z_{k,r}}(x)^{\frac{1}{w_k}}$. From the CDF of the SNR of user k when it is selected [49], the corresponding CDF of the throughput is

$$F_{Z_{k,r}|k^*=k}(x) = F_{Z_{k,r}}(x)^{\frac{\sum_{j=1}^{K} w_j}{w_k}} = F_{Z_{k,r}}(x)^{\frac{1}{w_k}}, \tag{6.4}$$

where $\sum_{j=1}^{K} w_j = 1$.

In opportunistic scheduling, users feed back the achievable throughput only when the weighted CDF of the throughput of a user k is greater than a certain threshold.

$$F_{Z_{k,r}}(x_k)^{\frac{1}{w_k}} \geq const = F_{Z_{k,r}}(\gamma_k^*)^{\frac{1}{w_k}} \quad \forall k = 1, \ldots, K, \tag{6.5}$$

In CDF based scheduling, we do not assign the same threshold for all the users but assign each user k a distinct threshold γ_k^*. These throughput thresholds of users are assigned so that their weighted CDF have the same value $F_{Z_{k,r}}(\gamma_k^*)^{\frac{1}{w_k}} = F_{Z_{j,r}}(\gamma_j^*)^{\frac{1}{w_j}}$, which then allows us to obtain $F_{Z_{j,r}}(\gamma_j^*) = F_{Z_{k,r}}(\gamma_k^*)^{\frac{w_j}{w_k}}$ for all user $j = 1, \ldots, K$. As a user k does not feed back with probability $Pr\{x_k < \gamma_k^*\} = F_{Z_{k,r}}(x_k)$, the threshold is selected so probability there are no user to feed back is negligible (smaller than ϵ).

$$\prod_{j=1}^{K} F_{Z_{j,r}}(\gamma_j^*) = F_{Z_{k,r}}(\gamma_k^*)^{\frac{1}{w_k}} \leq \epsilon. \tag{6.6}$$

Based on system requirement ϵ, the corresponding threshold can be found

$$\gamma_k^* = F_{Z_{k,r}}^{-1}(\epsilon^{w_k}) \ \forall k = 1, \ldots, K. \tag{6.7}$$

Lemma 1. *When the number of user K goes to infinity, there is an infinite number of users whose weights approach 0 and there is a finite number of users whose weights does not approach 0.*

Proof. We prove this lemma by contradiction.

Assume there is a set T_0 contains a finite number of users $K_0 = |T_0|$ whose weights approach 0, and a set T_1 contains an infinite number of users $K_1 = |T \backslash T_0| = K - K_0$ whose weights are greater than 0.

For users in T_1, there exists a value $w_{min} > 0$ such that all these weights are larger than w_{min}. Then, $\sum_{k=1}^{K} w_k \geq \sum_{j \in T_1} w_j \geq K_1 w_{min} = \infty$ which contradicts $\sum_{k=1}^{K} w_k = 1$.

Thus, K_1 is finite and hence, $K_0 = K - K_1$ is infinite.

□

Proposition 5. *The total amount of feedback in the system when the number of* *user* $K \to \infty$ *is*

$$N_{fb} = \frac{N}{\eta}\ln\frac{1}{\epsilon} + \frac{N}{\eta}\sum_{k\in T_1}(1-\epsilon^{w_k}+w_k\ln\epsilon),$$

where $0<\epsilon<1$ *is the probability there is no user to provide feedback.*

Proof. A user k only feeds back when its SNR is greater than the threshold SNR γ_k defined in equation (6.7). As probability that user k to feed back is $1-F_{Z_{k,r}}(\gamma_k^*)$, the average total number of feedback is

$$N_{fb} = \frac{N}{\eta}\sum_{k=1}^{K}\left(1-F_{Z_{k,r}}(\gamma_k^*)\right) = \frac{N}{\eta}\sum_{k=1}^{K}(1-\epsilon^{w_k}) \tag{6.8}$$

Because there are infinite number of users whose weights approach 0 when $K\to\infty$ as in Lemma 1, we consider two groups of users, one with constant weights and the other with weights approach 0. For any the user k with $w_k \to 0$, we can approximate $\epsilon^{w_k} = e^{w_k\ln\epsilon} = 1 + w_k\ln\epsilon + O(w_k^2)$. Then, we have

$$\begin{aligned} N_{fb} &= \frac{N}{\eta}\sum_{k=1}^{K}(1-\epsilon^{w_k}) \\ &= -\frac{N}{\eta}\sum_{k\in T_0} w_k\ln\epsilon + \frac{N}{\eta}\sum_{k\in T_1}(1-\epsilon^{w_k}) \\ &= \frac{N}{\eta}\ln\frac{1}{\epsilon} + \frac{N}{\eta}\sum_{k\in T_1}(1-\epsilon^{w_k}+w_k\ln\epsilon). \end{aligned}$$

In the extreme case where all the users have comparable weights or equivalently, there is no user with weight greater than 0, the total amount of feedback converges to a constant $\frac{N}{\eta}\ln\frac{1}{\epsilon}$. □

6.2.2 Best M feedback

Probability user k to feedback its CQI on a resource block r is $\frac{M\eta}{N}$. The BS upon receiving feedback from the users will select a user using the CDF based scheduling. Herein, A denotes the set of users to provide feedback and $1_A(k)$

denotes the presence of user k in A. The probability user k is selected is [49]

$$Pr\{k^* = k\} \\ = \sum_{\pi(A)} \frac{1_A(k) w_k}{\sum_{j=1}^{K} 1_A(j) w_j} \left(\frac{M\eta}{N}\right)^m \left(1 - \frac{M\eta}{N}\right)^{K-m}, \tag{6.9}$$

where $1_A(k)$ is an indicator function which shows if user k in the set of users to provide feedback on the considering resource block, $m = |A|$ is the total number of users to provide feedback. Where $\pi(A)$ represents all possible set of users to provide feedback.

To ensure negligible probability that there is no user to feed back, we choose M so that $\left(1 - \frac{M\eta}{N}\right)^K \leq \epsilon$. Equivalently, $M \geq \frac{N}{\eta}\left(1 - \epsilon^{\frac{1}{K}}\right)$.

We also set the constraint that each user feeds back CQI of at least one resource block. Thus, we have

$$M \geq \max\left(1, \frac{N}{\eta}\left(1 - \epsilon^{\frac{1}{K}}\right)\right). \tag{6.10}$$

Given ϵ, this equation allows us to choose the smallest M to minimize feedback requirement of the system.

Proposition 6. *The total number of feedback in best M feedback strategy when the number of user K goes to infinity is $N_{fb} = K$.*

Proof. As $\lim_{K\to\infty} \frac{N}{\eta}\left(1 - \epsilon^{\frac{1}{K}}\right) = 0$, combined with equation (6.10) we have $M = 1$. Thus, the total number of feedback is $N_{fb} = KM = K$. □

6.2.3 Hybrid feedback

In this scheme, we combine best M scheme and opportunistic feedback schemes. Each user select the best M among $\frac{N}{\eta}$ CQI on all of its sub-bands. Probability the CQI on a sub-band r is among the best M CQI of user k is $\frac{M\eta}{N}$. User k will feedback when the CQI x_k greater than the threshold γ_k^*. The probability a user k to provide feedback on this sub-band is $\frac{M\eta}{N}(1 - F_{Y_{k,r}}(\gamma_k^*))$.

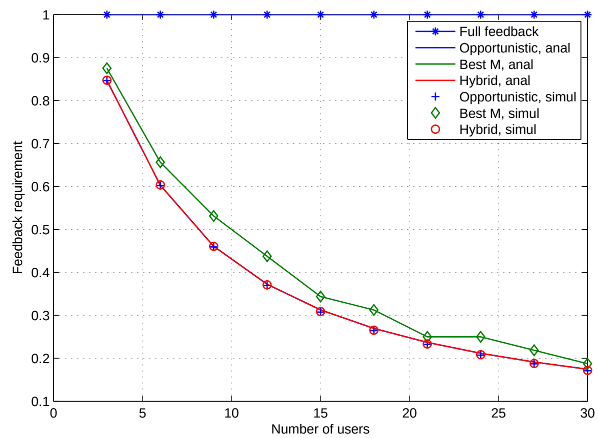


Figure 6.1: Normalized feedback requirement as a function of number of users, given $\epsilon = 10^{-3}$, $\eta = 1$.

Then, the average total feedback of the system is

$$N_{fb} = \sum_{j=1}^{K} M(1 - F_{Y_{k,r}}(\gamma_k^*)^{\frac{w_j}{w_k}}) \\ = KM - M \sum_{j=1}^{K} F_{Y_{k,r}}(\gamma_k^*)^{\frac{w_j}{w_k}}, \tag{6.11}$$

where $Y_{k,r}$ represents the throughput of the best M throughput on the sub-bands of user k.

We want to minimize the total system feedback while satisfying the constraint on the probability that there is no user to provide feedback. Probability a user k not to provide feedback is $1 - \frac{M\eta}{N}(1 - F_{Y_{k,r}}(\gamma_k^*))$. Then, we have the probability that there is no user to feed back must be smaller than ϵ

$$\prod_{j=1}^{K} \left(1 - \frac{M\eta}{N}(1 - F_{Y_{j,r}}(\gamma_k^*)^{\frac{w_j}{w_k}})\right) \leq \epsilon, \tag{6.12}$$

where the thresholds γ_k^* of users $k = 1, \ldots, K$ are chosen similar to equation (6.5). We then minimize the amount of feedback with the discussed constraints as in the following optimization problem

$$\begin{aligned} \min_{M^*,\gamma_k^*} \quad & KM - M\sum_{j=1}^{K} F_{Y_{k,r}}(\gamma_k^*)^{\frac{w_j}{w_k}} \\ s.t. \quad & 1 \le M \le \frac{N}{\eta} \\ & \prod_{j=1}^{K}\left(1 - \frac{M\eta}{N}(1 - F_{Y_{j,r}}(\gamma_k^*)^{\frac{w_j}{w_k}})\right) \le \epsilon \end{aligned} \tag{6.13}$$

Using an mixed integer nonlinear optimization solver [108], we find the optimal value M and γ_k^* for user $k = 1, \ldots, K$.

6.3 Simulation results

In the simulation, we consider a system with $N_c = 256$ subcarriers, where each $R_c = 8$ adjacent subcarriers are grouped into a resource block. Thus, there are $N = \frac{N_c}{R_c} = 32$ resource blocks. The number of taps in the delay spread channels is $L = 16$, $\delta = 4$, the transmit SNR $\rho = 10$ dB. The subbands are sizeable and consist of 1 to 4 resource blocks.

In figure 6.1, the feedback requirement normalized by full feedback as a function of number of users are shown. In this experiment, $\epsilon = 10^{-3}$ and each subband has only $\eta = 1$ resource blocks. When number of users increases, the full feedback requirement NK increases linearly with K. System feedback requirement of these schemes though increase but does increase with a slower rate in comparison to the full feedback case. The normalized feedback requirement shows the opportunistic and the hybrid schemes require least feedback. In our experiments, the optimization for parameters of the hybrid scheme results in the full feedback $M = N$ and the corresponding thresholds for all the users. In fact, the hybrid scheme is equivalent to the opportunistic scheme.

In figure 6.2, feedback requirement as a function of ϵ is shown. In this experiment, the number of users $K = 20$ and $\eta = 1$ are fixed. When ϵ increases, the system requires less feedback as the condition on the probability that there

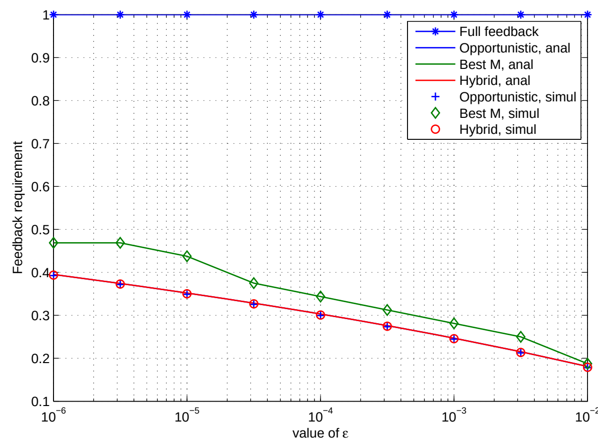


Figure 6.2: Normalized feedback requirement as a function of ϵ, given $K = 20$, $\eta = 1$.

is at least one feedback is easier to be satisfied. The hybrid and opportunistic schemes also outperform the optimized best M method. Unlike in [105] where the value M in best M method is fixed, we optimize value of M so the feedback requirement is minimized but still guarantee the probability there is no feedback at the BS is smaller than ϵ.

Average system throughput as a function of ϵ is considered in figure 6.3. We choose the hybrid scheme which is equivalent to the opportunistic scheme and is the optimal one to be investigated. The three curves correspond to three value of $\eta = 1$, 2 and 4. From the figure, the average system throughput slightly reduces when ϵ increases. This can be understood as the direct consequence of the intended reduction in the number of fed back CQI when ϵ increases. Moreover, when we group more resource blocks into a subband, the performance reduces. The reason is the rate of a subband is the average rate of η resource blocks. Using the average rate, we can not exploit the diversity in frequency dimension of these η resource blocks which causes the observed loss in system throughput as reflexed in figure 6.3 .

In figures 6.3 and 6.4, we consider the dependence of system throughput and the total feedback requirement on η. When η increases, average system throughput

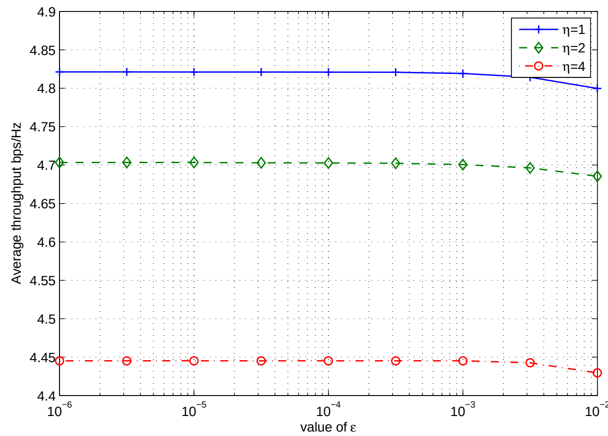


Figure 6.3: Hybrid scheme: average throughput as a function of ϵ, given $K = 20$.

decreases as more correlated resource blocks are grouped which decreases the quality of fed back information as well as the diversity in frequency dimension of the channel. Herein, the feedback requirement is reduced linearly with the reduction in the number of sub-bands. Furthermore, the CQI thresholding method helps further reduce system feedback requirement.

In figure 6.5, system throughput as a function of ϵ with different values of delay spread L is shown. When L increases, system throughput slightly decreases due to the loss in the capability of exploiting the diversity in frequency dimension of the channels of users. With different value of L, when ϵ increases, the feedback requirement reduces with the tradeoff in the decrement in system performance.

6.4 Conclusions

In this paper, we investigate performance of different feedback schemes which are the opportunistic, best M and hybrid scheme, in an OFDMA system applying CDF based scheduling. The threshold as well as the value of M are optimized based on the requirement on the probability of feedback availability at the BS on each resource block. Suitably selecting feedback parameters helps significantly reduce the overall feedback requirement of the system while the fairness

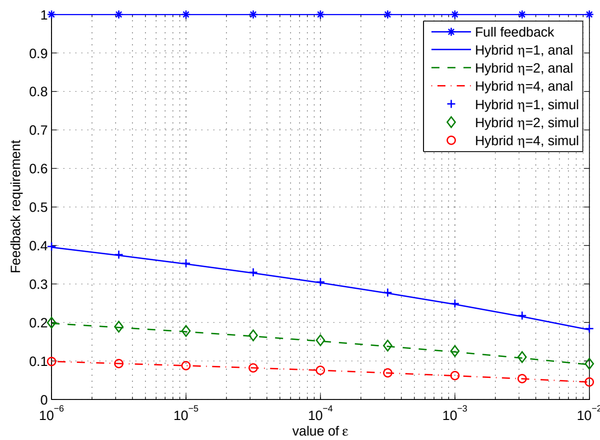


Figure 6.4: Hybrid scheme: normalized feedback requirement as a function of ϵ, given $K = 20$.

among users and good system throughput are still maintained. The analysis and simulation show when the number of users increases, or the requirement on feedback availability at the BS decreases, or the delay spread of the wideband channel decreases, feedback parameters can be suitably set so to further reduce the feedback requirement of the system. The simulation shows that the optimized hybrid scheme is identical to the optimized opportunistic scheme. Numerically, these two schemes can obtain higher system throughput than the best M scheme does.

The text of this chapter, in full, is a reprint of the paper [99], Anh H. Nguyen, Yichao Huang and Bhaskar D. Rao Novel partial feedback schemes and their evaluation in an OFDMA system with CDF based scheduling, that has been published in Asilomar Conference, 2013. The dissertation author is the primary researcher and author, and the co-authors listed in this publication collaborated and supervised the research constitutes this chapter.

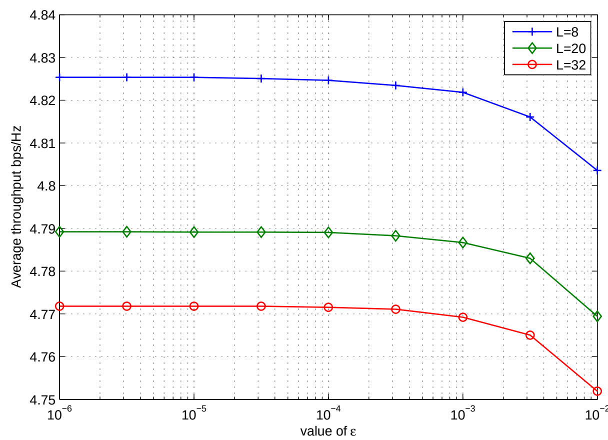


Figure 6.5: Hybrid scheme: throughput as a function of ϵ with different delay spread L, given $K = 20$, and $\eta = 1$.

Chapter 7

Conclusions

7.1 Conclusions

In this dissertation, we study the problem of multi-user diversity and fair resource allocation in wireless networks. In Chapter II, we studied multi-user diversity in wireless systems. As there are multiple users in the system with very different channel conditions, choosing a user with good channel condition is beneficial to overall system performance. Yet in a multiple-antennas system, user selection in one spatial mode affects the followed to be selected users as there is mutual interference between users as their channels are not perfectly orthogonal. The complexity grows with the size of the system, which is proportional to the number of users and number of antennas. To simplify the process, we developed the Greedy user selection algorithm to choose a reasonably good set of users for communication with the BS. We then proposed committee machine and evolutionary approaches to further improve the quality of the set of selected users. It is numerically shown the performance of our selection schemes approaches the performance of an exhaustive search. But because we truncate the search space in each step, my selection schemes have much lower complexity even with the committee machine and evolutionary approaches.

We have successfully developed classes of CDF scheduling algorithms which can be applied to real systems. The NPCS class of algorithms work efficiently without requiring prior knowledge of channel distribution. This means we do not need to worry about the transmission environment with various factors such as LoS/NLoS, surrounding objects, and so on. The algorithms remember channel information from the previous time interval and use this collected knowledge directly for scheduling purposes. In this type of algorithms, NPCS-2 is shown to be optimal in term of overall system throughput. Interestingly, these algorithms often compare favorably with a widely-used PF scheduling algorithm. When the transmission channel can be modeled, we propose using PCS algorithms, which effectively learn the parameters of the channel distribution functions and use these parameters to calculate the CDF scheduling metric.

Furthermore, CDF scheduling algorithms for discrete rate transmission are also developed. Because only discrete transmission rate is used in real systems, the throughput metric used in developing NPCS algorithms cannot be used directly in practical systems. Instead, the newly-developed algorithms incorporate a discrete transmission rate and limited feedback resources in their working procedures. The study shows that the algorithms can achieve the highest possible throughput and still satisfy the constrained feedback availability. Additionally, the algorithms can be customized for each individual user in order to maximally exploit the diversity in channel distribution of users.

We then analyze the performance of the CDF scheduling in an OFDMA system with either macro users, relay users, or both macro and relay users. In this system, partial feedback is assumed where each user feeds back only the best M among the CQI on all resource blocks. The study finds that applied partial feedback scheme alters the allocation for users. Because of this, we propose a remedy that adjusts users weights to guarantee proportional allocation for all users. When relays are incorporated into the system, the resource starvation phenomenon is observed due to the difference in coherent time on the two links between the BS and a relay, and between the relay and a user. To deal with resource starvation, we propose interpolating the CDF of user channels, which will help improve short-term fairness among users.

Though my study carefully investigates partial feedback in an OFDMA system, it is still unclear whether the feedback system is optimal or not. To address this question, I pursue a study on different feedback schemes for multi-user system applying CDF scheduling. The feedback schemes are thresholding, best M, and a combination of the two schemes, named the hybrid scheme. My numerical experiments show that among these schemes, the thresholding and the optimized hybrid scheme are identical. The objective is the overall feedback requirement,with a constraint on the feedback availability on each sub-carrier.

This thesis studies and solves the problem of fair resource allocation and multi-user diversity. We suggest exploiting multi-user diversity to maximize system performance. The first developed scheduling methods are able to exploit the channel distribution functions for a scheduling purpose. These algorithms are simple and able to control user fairness and priorities. They offer performance superiority in comparison to existing methods. We also developed methods so that the algorithms can be applied to real networks with more practical factors such as channel-coherent time, channel correlation, discrete rate transmission, and the availability of partial feedback.

7.2 Ongoing works

We recommend further research into how CDF scheduling works in temporally correlated channels. Intuitively, it makes sense that when a good channel is temporally correlated, it is likely to remain good in the next scheduling interval. Similarly, a bad channel remains bad. This causes a user with bad channel conditions to wait a long time until it can be allocated again. To prevent this resource starvation effect, I propose de-correlating users channels before scheduling. The proposed method has been shown to effectively prevent starvation with a small throughput tradeoff.

Another unanswered question is how CDF scheduling works in real systems where each user has QoS requirement. Until now, we have always assumed that each user has been assigned a suitable weight. In fact, users only give us requirements on throughput and QoS associated with the using applications. Future researchers need to develop ways to determine corresponding weight for the users. The assigned weights must capture the users throughput and QoS. The system must also define how to admit users so it can work stably.

Other possible topics can be listed as follows

1. Using CDF scheduling in a MIMO system is difficult. It is not easy to observe the distribution of SNR on a channel that corresponds to a spatial mode of a random channel matrix. Because of this, it seems that CDF scheduling is not ready to be applied on MIMO systems. It is even harder if the elements of the random channel matrix have different distribution, which depend on types of fading and the locations of users.

2. The performance of the developed algorithms needs to be evaluated in a changing environment. Channel statistics are not stable, but change over

time due to the users movement. A user can be served by a macro BS, or can enter the serving area of a smaller cell. In this case, it is desirable to find remedies that effectively switch the connections but still maintain users priority.

Bibliography

[1] D. Park and B. G. Lee, "Qos support by using cdf-based wireless packet scheduling in fading channels," *IEEE Trans. on Commun.*, vol. 54, no. 11, pp. 2051–2061, Nov. 2006.

[2] R. Knopp and P. Humblet, "Information capacity and power control in single-cell multiuser communications," in *IEEE Int. Conf. on Commun. (ICC)*, vol. 1, Jun. 1995.

[3] E. Hahne, "Round-robin scheduling for max-min fairness in data networks," *IEEE J. on Sel. Areas in Commun.*, vol. 9, no. 7, pp. 1024–1039, Sep. 1991.

[4] X. Liu, E. K. P. Chong, and N. Shroff, "Opportunistic transmission scheduling with resource-sharing constraints in wireless networks," *IEEE J. on Sel. Areas in Commun.*, vol. 19, no. 10, pp. 2053–2064, Oct. 2001.

[5] F. P. Kelly, A. K. Maulloo, and D. K. Tan, "Rate control for communication networks: shadow prices, proportional fairness and stability," *J. of the Operational Research society*, vol. 49, no. 3, pp. 237–252, 1998.

[6] A. Jalali, R. Padovani, and R. Pankaj, "Data throughput of CDMA-HDR a high efficiency-high data rate personal communication wireless system," in *IEEE Veh. Tech. Conf.*, vol. 3, 2000, pp. 1854–1858.

[7] C. Wengerter, J. Ohlhorst, and A. von Elbwart, "Fairness and throughput analysis for generalized proportional fair frequency scheduling in ofdma," in *IEEE Veh. Tech. Conf.*, vol. 3, 2005, pp. 1903–1907.

[8] R. La and V. Anantharam, "Utility-based rate control in the internet for elastic traffic," *IEEE/ACM Trans. on Networking*, vol. 10, no. 2, pp. 272–286, Apr. 2002.

[9] B. G. Lee, D. Park, and H. Seo, *Wireless communications resource management.* Wiley. com, 2009.

[10] G. Caire and S. Shamai, "On the achievable throughput of a multiantenna gaussian broadcast channel," *IEEE Trans. on Inf. Theory*, vol. 49, no. 7, pp. 1691–1706, Jul. 2003.

[11] N. Jindal, W. Rhee, S. Vishwanath, S. Jafar, and A. Goldsmith, "Sum power iterative water-filling for multi-antenna gaussian broadcast channels," *IEEE Trans. on Inf. Theory*, vol. 51, no. 4, pp. 1570–1580, Apr. 2005.

[12] B. Song, R. Cruz, and B. Rao, "Network duality for multiuser MIMO beamforming networks and applications," *IEEE Trans. on Commun.*, vol. 55, no. 3, pp. 618–630, Mar. 2007.

[13] W. Rhee, W. Yu, and J. Cioffi, "The optimality of beamforming in uplink multiuser wireless systems," *IEEE Trans. on Wireless Commun.*, vol. 3, no. 1, pp. 86–96, Jan. 2004.

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

[15] P. Lu and H.-C. Yang, "Sum-rate analysis of multiuser MIMO system with zero-forcing transmit beamforming," *IEEE Trans. on Commun.*, vol. 57, no. 9, pp. 2585–2589, Sep. 2009.

[16] G. Ginis and J. Cioffi, "A multi-user precoding scheme achieving crosstalk cancellation with application to DSL systems," in *Asilomar Conf.*, vol. 2, 2000, pp. 1627–1631.

[17] G. H. Golub and C. F. van Van Loan, *Matrix computations.* The Johns Hopkins University Press, 1996.

[18] C. D. Meyer, Jr, "Generalized inverses of triangular matrices," *SIAM Journal on Applied Mathematics*, vol. 18, no. 2, pp. 401–406, 1970.

[19] S. Haykin, *Neural networks: a comprehensive foundation.* Prentice Hall PTR, 1999.

[20] A. H. Nguyen and B. D. Rao, "User selection schemes for maximizing throughput of multiuser MIMO systems using zero forcing beamforming," in *ICASSP*, 2011.

[21] D. Park, H. Seo, H. Kwon, and B. G. Lee, "Wireless packet scheduling based on the cumulative distribution function of user transmission rates," *IEEE Trans. on Commun.*, vol. 53, no. 11, pp. 1919 – 1929, Nov. 2005.

[22] M. Sharif and B. Hassibi, "On the capacity of MIMO broadcast channels with partial side information," *IEEE Trans. on Inf. Theory*, vol. 51, no. 2, pp. 506–522, Feb. 2005.

[23] D. Love, R. Heath, V. Lau, D. Gesbert, B. Rao, and M. Andrews, "An overview of limited feedback in wireless communication systems," *IEEE J. on Sel. Areas in Commun.*, vol. 26, no. 8, pp. 1341 –1365, Oct. 2008.

[24] A. Ghosh, R. Ratasuk, B. Mondal, N. Mangalvedhe, and T. Thomas, "LTE-advanced: next-generation wireless broadband technology [invited paper]," *IEEE Wireless Commun.*, vol. 17, no. 3, pp. 10 –22, Jun. 2010.

[25] S. Sanayei and A. Nosratinia, "Opportunistic downlink transmission with limited feedback," *IEEE Trans. on Inf. Theory*, vol. 53, no. 11, pp. 4363 –4372, Nov. 2007.

[26] L. Massoulie and J. Roberts, "Bandwidth sharing: objectives and algorithms," *IEEE/ACM Trans. on Networking*, vol. 10, no. 3, pp. 320 –328, Jun. 2002.

[27] H. Fattah and C. Leung, "An overview of scheduling algorithms in wireless multimedia networks," *IEEE Wireless Commun.*, vol. 9, no. 5, pp. 76–83, Dec. 2002.

[28] A. Asadi and V. Mancuso, "A survey on opportunistic scheduling in wireless communications," *IEEE Commun. Surveys Tutorials*, pp. 1–18, Jan. 2013.

[29] D. Wu and R. Negi, "Downlink scheduling in a cellular network for quality-of-service assurance," *IEEE Trans. on Veh. Tech.*, vol. 53, no. 5, pp. 1547 –1557, Sep. 2004.

[30] Z. Han, Z. Ji, and K. Liu, "Fair multiuser channel allocation for OFDMA networks using nash bargaining solutions and coalitions," *IEEE Trans. on Commun.*, vol. 53, no. 8, pp. 1366–1376, Aug. 2005.

[31] H. Boche and M. Schubert, "Nash bargaining and proportional fairness for wireless systems," *IEEE/ACM Trans. on Networking,*, vol. 17, no. 5, pp. 1453–1466, Oct. 2009.

[32] Z. Shen, J. Andrews, and B. Evans, "Adaptive resource allocation in multiuser OFDM systems with proportional rate constraints," *IEEE Trans. on Wireless Commun.*, vol. 4, no. 6, pp. 2726–2737, Nov. 2005.

[33] H. Huh, S.-H. Moon, Y.-T. Kim, I. Lee, and G. Caire, "Multi-cell MIMO downlink with cell cooperation and fair scheduling: A large-system limit analysis," *IEEE Trans. on Inf. Theory*, vol. 57, no. 12, pp. 7771–7786, Dec. 2011.

[34] M. Salem, A. Adinoyi, M. Rahman, H. Yanikomeroglu, D. Falconer, and Y.-D. Kim, "Fairness-aware radio resource management in downlink OFDMA cellular relay networks," *IEEE Trans. on Wireless Commun.*, vol. 9, no. 5, pp. 1628–1639, May 2010.

[35] J. Jaramillo and R. Srikant, "Optimal scheduling for fair resource allocation in ad hoc networks with elastic and inelastic traffic," *IEEE/ACM Trans. on Networking*, vol. 19, no. 4, pp. 1125–1136, Aug. 2011.

[36] M. Ergen, S. Coleri, and P. Varaiya, "Qos aware adaptive resource allocation techniques for fair scheduling in OFDMA based broadband wireless access systems," *IEEE Trans. on Broadcasting,*, vol. 49, no. 4, pp. 362–370, Dec. 2003.

[37] G. Song and Y. Li, "Cross-layer optimization for OFDM wireless networks-part II: algorithm development," *IEEE Trans. on Wireless Commun.*, vol. 4, no. 2, pp. 625–634, Mar. 2005.

[38] J.-G. Choi and S. Bahk, "Cell-throughput analysis of the proportional fair scheduler in the single-cell environment," *IEEE Trans. on Veh. Tech.*, vol. 56, no. 2, pp. 766 –778, Mar. 2007.

[39] N. Sharma and L. Ozarow, "A study of opportunism for multiple-antenna systems," *IEEE Trans. on Inf. Theory*, vol. 51, no. 5, pp. 1804 – 1814, May. 2005.

[40] H. Kwon, Y. Kim, J.-K. Han, D. Kim, H. W. Lee, and Y. K. Kim, "Performance evaluation of high-speed packet enhancement on CDMA2000 1xEV-DV," *IEEE Comm. Mag.*, vol. 43, no. 4, pp. 67–73, Apr. 2005.

[41] S. Sigdel and W. Krzymien, "Simplified fair scheduling and antenna selection algorithms for multiuser MIMO orthogonal space-division multiplexing downlink," *IEEE Trans. on Veh. Tech.*, vol. 58, no. 3, pp. 1329–1344, Mar. 2009.

[42] R. Kwan, C. Leung, and J. Zhang, "Proportional fair multiuser scheduling in LTE," *IEEE Signal Process/ Letters*, vol. 16, no. 6, pp. 461–464, Jun. 2009.

[43] Z. Shen, J. G. Andrews, and B. L. Evans, "Adaptive resource allocation in multiuser OFDM systems with proportional rate constraints," *IEEE Trans. on Wireless Commun.*, vol. 4, no. 6, pp. 2726–2737, Nov. 2005.

[44] H. J. Kushner and P. A. Whiting, "Convergence of proportional-fair sharing algorithms under general conditions," *IEEE Trans. on Wireless Commun.*, vol. 3, no. 4, pp. 1250–1259, Jul. 2004.

[45] P. Svedman, S. K. Wilson, L. J. Cimini, and B. Ottersten, "Opportunistic beamforming and scheduling for OFDMA systems," *IEEE Trans. on Commun.*, vol. 55, no. 5, pp. 941 –952, May. 2007.

[46] G. Miao, N. Himayat, Y. G. Li, and A. Swami, "Cross-layer optimization for energy-efficient wireless communications: a survey," *Wireless Commun. and Mobile Comp.*, vol. 9, no. 4, pp. 529–542, Apr. 2009.

[47] M. Andrews, "Instability of the proportional fair scheduling algorithm for HDR," *IEEE Trans. on Wireless Commun.*, vol. 3, no. 5, pp. 1422–1426, Sep. 2004.

[48] Y. Huang and B. D. Rao, "An analytical framework for heterogeneous partial feedback design in heterogeneous multicell OFDMA networks," *IEEE Trans. on Signal Proc.*, vol. 61, no. 3, pp. 753–769, Feb. 2013.

[49] A. Nguyen, Y. Huang, and B. D. Rao, "Weighted CDF-based scheduling for an OFDMA relay downlink with partial feedback," in *IEEE Asilomar Conf. on Signals, Systems and Computers*, Nov. 2012.

[50] Y. Huang and B. Rao, "Random beamforming with heterogeneous users and selective feedback: Individual sum rate and individual scaling laws," *IEEE Trans. on Wireless Commun.*, vol. 12, no. 5, pp. 2080–2090, May. 2013.

[51] S. Patil and G. De Veciana, "Measurement-based opportunistic scheduling for heterogenous wireless systems," *IEEE Trans. on Commun.*, vol. 57, no. 9, pp. 2745–2753, 2009.

[52] M. Kountouris, T. Sälzer, and D. Gesbert, "Scheduling for multiuser MIMO downlink channels with ranking-based feedback," *EURASIP J. on Advances in Signal Proc.*, p. 131, 2008.

[53] H. A. David and H. N. Nagaraja, *Order Statistics, Third Edition.* Wiley, 2003.

[54] I. S. Gradshteyn and I. M. Ryzhik, *Table of integrals, series, and products*, 7th ed., 2007.

[55] A. Gersho and R. M. Gray, *Vector quantization and signal compression.* Kluwer Academic Publishers, 1991.

[56] T. K. Moon and W. C. Stirling, *Mathematical methods and algorithms for signal processing*, 2000, vol. 1.

[57] N. Beaulieu and C. Cheng, "Efficient Nakagami-m fading channel simulation," *IEEE Trans. on Veh. Tech.*, vol. 54, no. 2, pp. 413–424, Mar. 2005.

[58] A. Nguyen, Y. Huang, and B. Rao, "Learning methods for CDF scheduling in multiuser heterogeneous systems," *IEEE Trans. on Signal Proc.*, 2013, submitted.

[59] Z. Fang and B. Bensaou, “Fair bandwidth sharing algorithms based on game theory frameworks for wireless ad-hoc networks,” in *INFOCOM*, vol. 2. IEEE, 2004, pp. 1284–1295.

[60] Z. Jiang, Y. Ge, and Y. Li, “Max-utility wireless resource management for best-effort traffic,” *IEEE Trans. on Wireless Commun.*, vol. 4, no. 1, pp. 100–111, Jan. 2005.

[61] S. Stanczak, M. Wiczanowski, and H. Boche, “Distributed utility-based power control: Objectives and algorithms,” *IEEE Trans. on Signal Proc.*, vol. 55, no. 10, pp. 5058–5068, Oct. 2007.

[62] D. P. Bertsekas and R. G. Gallager, *Data Networks*, 2nd ed. Prentice Hall, 1992.

[63] L. Tassiulas and S. Sarkar, “Maxmin fair scheduling in wireless networks,” in *INFOCOM*, vol. 2, 2002, pp. 763–772.

[64] D. Gesbert and M. S. Alouini, “How much feedback is multi-user diversity really worth?” in *Proc. IEEE ICC*, 2004, pp. 234–238.

[65] N. Ravindran and N. Jindal, “Multi-user diversity vs. accurate channel feedback for mimo broadcast channels,” in *IEEE Int. Conf. on Commun.* IEEE, 2008, pp. 3684–3688.

[66] Y. Isukapalli and B. Rao, “Packet error probability of a transmit beamforming system with imperfect feedback,” *IEEE Trans. on Signal Proc.*, vol. 58, no. 4, pp. 2298 –2314, Apr. 2010.

[67] A. Kuhne and A. Klein, “Throughput analysis of multi-user OFDMA-systems using imperfect CQI feedback and diversity techniques,” *IEEE Journal on Sel. Areas in Commun.*, vol. 26, no. 8, pp. 1440 –1450, Oct. 2008.

[68] Y. Huang and B. Rao, “Performance analysis of heterogeneous feedback design in an OFDMA downlink with partial and imperfect feedback,” *IEEE Trans. on Signal Proc.*, vol. 61, no. 4, pp. 1033 –1046, Feb. 2013.

[69] J. Qi and S. Aissa, “Cross-layer design for multiuser MIMO MRC systems with feedback constraints,” *IEEE Trans. on Veh. Tech.*, vol. 58, no. 7, pp. 3347 –3360, Sept. 2009.

[70] T. Kim and M. Skoglund, “On the expected rate of slowly fading channels with quantized side information,” *IEEE Trans. on Commun.*, vol. 55, no. 4, pp. 820 –829, Apr. 2007.

[71] V. Hassel, D. Gesbert, M.-S. Alouini, and G. Oien, “A threshold-based channel state feedback algorithm for modern cellular systems,” *IEEE Trans. on Wireless Commun.*, vol. 6, no. 7, pp. 2422 –2426, Jul. 2007.

[72] J. Leinonen, J. Hamalainen, and M. Juntti, “Performance analysis of downlink OFDMA resource allocation with limited feedback,” *IEEE Trans. on Wireless Commun.*, vol. 8, no. 6, pp. 2927 –2937, Jun. 2009.

[73] V. Lau, W. K. Ng, and D. Hui, “Asymptotic tradeoff between cross-layer goodput gain and outage diversity in OFDMA systems with slow fading and delayed CSIT,” *IEEE Trans. on Wireless Commun.*, vol. 7, no. 7, pp. 2732 –2739, Jul. 2008.

[74] S. Donthi and N. Mehta, “Joint performance analysis of channel quality indicator feedback schemes and frequency-domain scheduling for LTE,” *IEEE Trans. on Veh. Tech.*, vol. 60, no. 7, pp. 3096–3109, Sep. 2011.

[75] A. Nguyen, Y. Huang, and B. D. Rao, “Performance of a multiuser downlink system applying thresholding feedback with imperfect channel information,” in *IEEE Veh. Tech. Conf.* IEEE, 2013.

[76] H. Stark and J. W. Woods, *Probability and Random Processes With Applications to Signal Processing*, 3rd ed. Prentice Hall, 2001.

[77] A. Nguyen, , and B. Rao, “Cdf scheduling algorithms for finite rate multiuser systems with limited feedback,” *IEEE Trans. on Wireless Commun.*, 2014, submitted.

[78] Q. Spencer, C. Peel, A. Swindlehurst, and M. Haardt, “An introduction to the multi-user MIMO downlink,” *IEEE Commun. Mag.*, vol. 42, no. 10, pp. 60 – 67, Oct. 2004.

[79] W. Rhee and J. Cioffi, “Increase in capacity of multiuser OFDM system using dynamic subchannel allocation,” in *IEEE Conf. Proc. on Veh. Tech.*, vol. 2, May. 2000, pp. 1085 –1089 vol.2.

[80] G. Dimic and N. Sidiropoulos, “On downlink beamforming with greedy user selection: performance analysis and a simple new algorithm,” *IEEE Trans. on Signal Proc.*, vol. 53, no. 10, pp. 3857 – 3868, Oct. 2005.

[81] P. Viswanath, D. Tse, and R. Laroia, “Opportunistic beamforming using dumb antennas,” *IEEE Trans. on Inf. Theory*, vol. 48, no. 6, pp. 1277 – 1294, Jun. 2002.

[82] F. Kelly, “Charging and rate control for elastic traffic,” *European trans. on Telecommun.*, vol. 8, no. 1, pp. 33–37, 1997.

[83] H. Shirani-Mehr and G. Caire, “Channel state feedback schemes for multiuser MIMO-OFDM downlink,” *IEEE Trans. on Commun.*, vol. 57, no. 9, pp. 2713 –2723, Sep. 2009.

[84] S.-H. Hur and B. Rao, "Sum rate analysis of a reduced feedback OFDMA downlink system employing joint scheduling and diversity," *IEEE Trans. on Signal Proc.*, vol. 60, no. 2, pp. 862 –876, Feb. 2012.

[85] J. Chen, R. Berry, and M. Honig, "Limited feedback schemes for downlink OFDMA based on sub-channel groups," *IEEE J. on Sel. Areas in Commun.*, vol. 26, no. 8, pp. 1451 –1461, Oct. 2008.

[86] H. Bolcskei, R. Nabar, O. Oyman, and A. Paulraj, "Capacity scaling laws in MIMO relay networks," *IEEE Trans. on Wireless Commun.*, vol. 5, no. 6, pp. 1433–1444, Jun. 2006.

[87] Y. Jing and H. Jafarkhani, "Single and multiple relay selection schemes and their achievable diversity orders," *IEEE Trans. on Wireless Commun.*, vol. 8, no. 3, pp. 1414 –1423, Mar. 2009.

[88] T. C.-Y. Ng and W. Yu, "Joint optimization of relay strategies and resource allocations in cooperative cellular networks," *IEEE J. on Sel. Areas in Commun.*, vol. 25, no. 2, pp. 328 –339, Feb. 2007.

[89] B. Zafar, S. Gherekhloo, and M. Haardt, "Analysis of multihop relaying networks: Communication between range-limited and cooperative nodes," *IEEE Veh. Tech. Mag.*, vol. 7, no. 3, pp. 40–47, Sep. 2012.

[90] C. Lo, S. Vishwanath, and R. Heath, "Relay subset selection in wireless networks using partial decode-and-forward transmission," *IEEE Trans. on Veh. Tech.*, vol. 58, no. 2, pp. 692 –704, Feb. 2009.

[91] A. Nosratinia, T. Hunter, and A. Hedayat, "Cooperative communication in wireless networks," *IEEE Commun. Magazine*, vol. 42, no. 10, pp. 74–80, Oct. 2004.

[92] R. Nabar, H. Bolcskei, and F. Kneubuhler, "Fading relay channels: performance limits and space-time signal design," *IEEE J. on Sel. Areas in Commun.*, vol. 22, no. 6, pp. 1099–1109, Aug. 2004.

[93] B. Rankov and A. Wittneben, "Spectral efficient protocols for half-duplex fading relay channels," *IEEE J. on Sel. Areas in Commun.*, vol. 25, no. 2, pp. 379–389, Feb. 2007.

[94] B. Wang, J. Zhang, and A. Host-Madsen, "On the capacity of MIMO relay channels," *IEEE Trans. on Inf. Theory*, vol. 51, no. 1, pp. 29–43, Jan. 2005.

[95] M. Hasna and M.-S. Alouini, "Optimal power allocation for relayed transmissions over rayleigh-fading channels," *IEEE Trans. on Wireless Commun.*, vol. 3, no. 6, pp. 1999–2004, Nov. 2004.

[96] G. Senarath, W. Tong, P. Zhu, H. Zhang, D. Steer, D. Yu, M. Naden, and D. Kitchener, "Multi-hop relay system evaluation methodology (channel model and performance metric)," *IEEE C802. 16j-06/013r3*, 2007.

[97] I. S. Gradshteyn and I. M. Ryzhik, *Table of integrals, series, and products*, 7th ed., 2007.

[98] *Numerical linear algebra and optimization*, 2011.

[99] A. Nguyen, Y. Huang, and B. D. Rao, "Novel partial feedback schemes and their evaluation in an OFDMA system with CDF based scheduling," in *Asilomar*, 2013.

[100] S. Peters, A. Panah, K. Truong, and R. Heath, "Relay architectures for 3GPP LTE-Advanced," *EURASIP J. on Wireless Commun. and Networking*, vol. 2009, no. 1, 2009.

[101] A. Nguyen, Y. Huang, and D. B. Rao, "Weighted CDF-based scheduling for an OFDMA relay downlink with partial feedback," *IEEE Trans. on Commun.*, submitted.

[102] 3GPP36.826, "Evolved universal terrestrial radio access (E-UTRA); relay radio transmission and reception," Mar. 2012.

[103] I. Wong and B. Evans, "Optimal resource allocation in the OFDMA downlink with imperfect channel knowledge," *IEEE Trans. on Commun.*, vol. 57, no. 1, pp. 232 –241, Jan. 2009.

[104] R. Agarwal, V. Majjigi, Z. Han, R. Vannithamby, and J. Cioffi, "Low complexity resource allocation with opportunistic feedback over downlink OFDMA networks," *IEEE J. on Sel. Areas in Commun.*, vol. 26, no. 8, pp. 1462 –1472, Oct. 2008.

[105] Y.-J. Choi and S. Bahk, "Partial channel feedback schemes maximizing overall efficiency in wireless networks," *IEEE Trans. on Wireless Commun.*, vol. 7, no. 4, pp. 1306 –1314, Apr. 2008.

[106] C. Diao, M. Chen, and B. Wu, "Comments on "partial channel feedback schemes maximizing overall efficiency in wireless networks"," *IEEE Trans. on Wireless Commun.*, vol. 10, no. 3, pp. 1006 –1008, Mar. 2011.

[107] Y. Al-Harthi, A. Tewfik, and M.-S. Alouini, "Multiuser diversity with quantized feedback," *IEEE Trans. on Wireless Commun.*, vol. 6, no. 1, pp. 330 –337, Jan. 2007.

[108] G. Sierksma, P. van Dam, and G. Tijssen, *Linear and integer programming: theory and practice.*