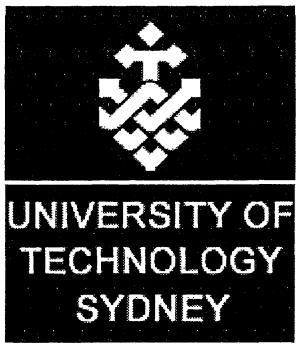
A Dissertation submitted in fulfillment of
the requirements for the degree of Doctor of Philosophy

Link Adaptation with Limited Feedback for Future Wireless Networks

Rachod Patachaianand

Autumn 2012



University of Technology, Sydney
Faculty of Engineering and Information Technology
Centre for Real Time Information Networks

Supervisor:

Dr. Kumbesan Sandrasegaran

Associate Professor,
Faculty of Engineering and Information Technology,
Centre for Real Time Information Networks,
University of Technology Sydney

UNIVERSITY OF TECHNOLOGY SYDNEY

Certificate of Originality

I certify that the work in this thesis has not previously been submitted for a degree nor has it been submitted as part of requirements for a degree except as fully acknowledged within the text.

I certify that the thesis has been written by me. Any help that I have received in my research work and the preparation of the thesis itself has been acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

Signature of Candidate

Production Note:
Signature removed prior to publication.

Acknowledgement

I would like to thank Dr. Kumbesan Sandrasegaran, my supervisor, for the great support in both general and academic aspects he has given me throughout the years since my first semester at UTS in Master of Engineering Studies in Autumn 2005.

I would like to thank my friends in CRIN laboratory who reviewed drafts of my work and discussed some technical aspects.

Lastly, I would like to give grateful thanks to my parents, Ampol and Sunee, who support and encourage me to achieve as high education as I can, and my sister Rasana who helps me while studying in Australia.

Abstract

Link adaptation technique (LAT) is one of the radio resource management (RRM) functions necessary for future wireless network to enhance the system capacity and provide an adequate quality of service to the end users.

LAT requires knowledge of users' received singal-to-noise ratio so that scheduling and adaptive modulation and coding scheme (MCS) can be performed to optimise the system performance. The received SNR is measured by the mobile users in the downlink transmission and converted to channel state information (CSI), which is sent via feedback channel in the uplink direction to the base station. The amount of feedback required for adaptive MCS and scheduling at the BS increases with the number of users which consumes significant amount of system resources. It is critical to provide an efficient feedback algorithm for LAT to achieve high utilisation of the system resources while maintaining the system performance close to the system with perfect knowledge of CSI.

This thesis attempted to answer the following questions related to LAT with reduced CSI feedback load:

a) Based on our current knowledge of feedback reduction techniques, can the feedback load be further reduced without scarifying the system performance?
b) Do multiple threshold schemes perform better than single threshold schemes?
c) How is it possible to develop a feedback technique(s) that can self-adjust the parameters associated with CSI reporting to suit the system conditions?
d) Provided that a well-known scheduling technique creates significant delays in the scheduling process, is it possible to mitigate the delays without any need for additional feedback resources?
e) Based on our current knowledge of CSI transmission on the uplink, can a more efficient media access control scheme for CSI reporting be developed?

UNIVERSITY OF TECHNOLOGY SYDNEY

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1. Brief History of Cellular Systems

During the last two decades, wireless cellular communication systems have dramatically grown. The first era of wireless cellular communication is referred to as the first-generation (1G) system. The 1G system was based on analog transmission and could offer only speech services. The multiple access technique used in 1G was very simple. This multiple access technology is referred to as Frequency Division Multiple Access (FDMA). Each user connecting to a 1G system was allocated a narrow/non-overlapping frequency band, and the user holds this band for the entire call period.

The significant improvement on the second generation (2G) system was the use of digital transmission technologies which allow error correction coding and more efficient multiple access technique such as Time Division Multiple Access (TDMA) to be used. The available frequency band in 2G was divided into several sub-channels and, with the use of TDMA techniques, many users can share one sub-channel on different timeslot. A good example of 2G systems is the GSM system which has been successful globally.

Universal Mobile Telecommunication System (UMTS) is one of the third generation (3G) technologies. It has been standardised by the 3rd Generation Partnership Project (3GPP) to provide higher data-rate applications which 2G technologies (i.e. GSM) could not support. Initially, the UMTS standard was designed to support a theoretical maximum composite data rate of up to 2 Mbps [1]. In practice and in ideal conditions, UMTS users could attain data rates up to 384 kbps which is much greater than 14.4 kbps provided by the earlier GSM technology. The maximum data rate was enhanced to 14.0 Mbps with the use of

High Speed Downlink Packet Access (HSDPA) in the downlink direction (base station to mobiles) [2] and to beyond 5 Mbps with the use of High Speed Uplink Packet Access (HSUPA) in the uplink direction (mobiles to base station) [3]. The achievable theoretical data rate of HSDPA has further increased to 42 Mbps with the simultaneously use of multi-antenna technique and higher-order modulation scheme.

There was a significant growth in the number of 3G/UMTS subscribers worldwide which has led to a significant growth in UMTS networks globally. Many GSM service providers have upgraded their GSM systems to UMTS. It is expected that the 3G/UMTS networks may reach their maximum capacity in the near future as its capability to provide high-speed services such as video on demand is limited, a further evolution in mobile telecommunication technology has been proposed by 3GPP.

The next 3GPP development to enhance mobile telecommunication networks is the new version of terrestrial radio access network, also called Long Term Evolution (LTE) [4][5]. It will result in the new evolved Release 8 of the UMTS standard including substantial extensions and modifications of the UMTS terrestrial radio access network (UTRAN) [6][7]. One of the major modifications of LTE compared with UMTS is the change in the air interface technology, from WCDMA to Orthogonal Frequency Division Multiple Access (OFDMA). In OFDMA based system, all users share the radio resource which comprises a number of clusters of subcarriers. This allows the system efficiency to be substantially improved by dynamic scheduling and link adaptation techniques.

Another significant modification in the radio access network is the introduction of Evolved UTRAN, which can support huge amount of data traffic while being compatible with legacy technologies such as UMTS. The first draft version of this standardisation was released in 2007. Based on the first draft-version of the LTE standardisation, research is being conducted throughout the world on finding techniques to optimise the usage of the radio resources in LTE. The evolution of the radio technologies is summarised in Figure 1.1.

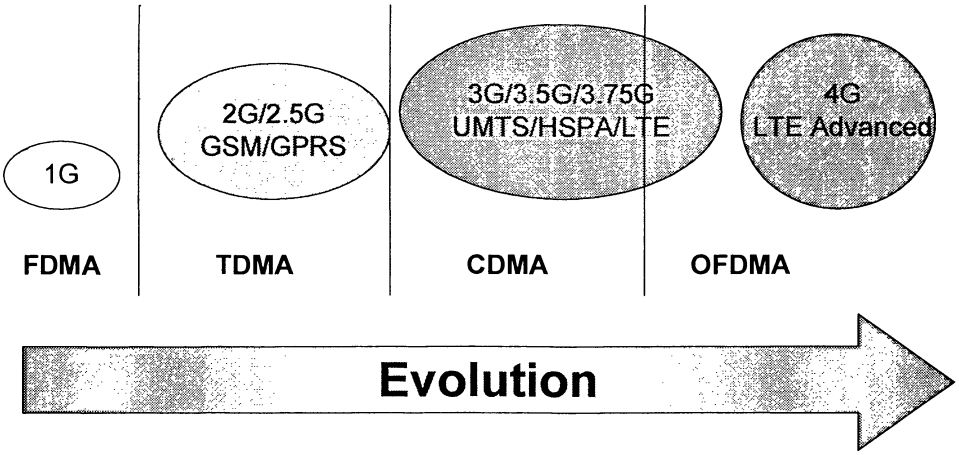


Figure 1.1 Evolution of radio access technologies

The key features of LTE are as following:

- all-IP flat architecture system.
- High theoretical download rates of up to 299.6 Mbit/s
- High theoretical upload rates up to 75.4 Mbit/s
- User of multi antenna technology (MIMO)
- Different user terminal classes defined from various applications.
- High system bandwidth of up to 20 MHz
- Low data transmission and handover latencies
- Low connection setup time
- Enhanced support for mobility,
- Advanced modulation techniques OFDMA for the downlink, SC-FDMA for the uplink
- Support for both FDD and TDD communication systems
- Simplified architecture: only of eNode Bs in EUTRAN
- Support for inter-operation and co-existence with legacy standards (e.g. GSM/EDGE, UMTS and CDMA2000).

1.2. Overview of Link Adaptation Techniques in Wireless Networks

In a wireless system, the received signal to noise ratio (SNR) at the receiver varies due to many factors such as noise, multipath fading effects and interference. If the

transmission parameters were set based on the worst case channel state, large amount of capacity would be unnecessarily wasted when the mobile experienced excellent channel conditions. Figure 1.2 shows a typical radio fading channel with fast fading.

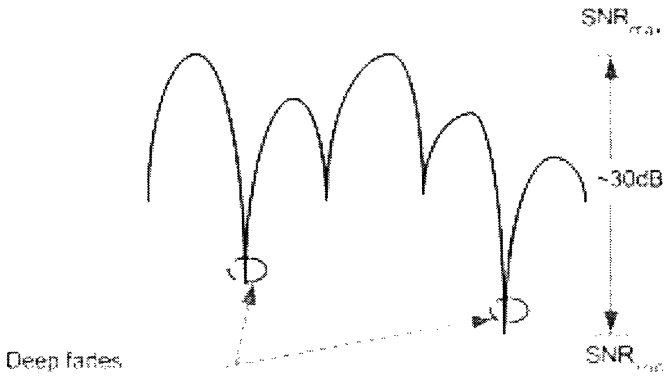


Figure 1.2 Typical radio fading channel.

From Figure 1.2, it can be observed that the channel may vary in order of 30 dB from the deep fading to the peak channel gain. If the system was designed to operate at this worst-case channel condition during deep fades, the channel utilisation would be inefficient.

In modern wireless networks where dynamic channel-sensitive link adaptation and resource allocation are implemented, the channel opportunistic scheduling (OpS), link adaptation technique (LAT), channel state information (CSI) reporting and Hybrid Automatic Repeat-Request (HARQ) must closely interact so that the system performance is maximised. The interconnection between OpS, LAT, HARQ and CSI is demonstrated in Figure 1.3.

In Figure 1.3, receivers send CSI through feedback channel. The transmitter then schedules the best users based on OpS policy. It determines the most optimal modulation and coding scheme (MCS) accordingly based on LAT criteria to suit the received SNR of the scheduled users before the data transmission takes place. The receivers perform HARQ and send ACK message whether the data is received correctly. Each function in Figure 1.3 will be discussed in more details in the next sections.

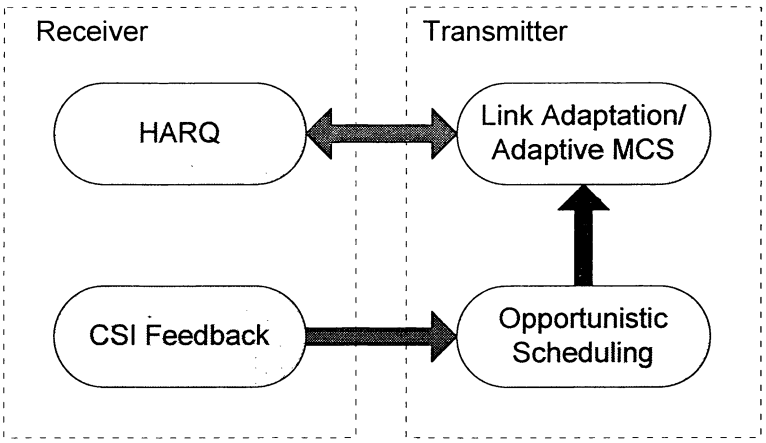


Figure 1.3 Interconnection of link adaptation, scheduling, HARQ and CSI feedback

1.2.1 Power Control and Rate Control

One of the techniques to avoid this undesirable situation is to apply an advanced link adaptation technique (LAT) to the system.

In WCDMA-based UMTS networks, the link adaptation technique is performed by means of fast power control (fast PC), at a basis of 1500 Hz, to compensate for the rapid changes in radio channel conditions. This mechanism is to maintain the radio link qualities at the receiver (mobile users in downlink communications) at approximately the same average SNR despite differences in user locations and channel conditions, see Figure 1.4 (a) below.

In HSDPA, the adaptation technique is performed by means of rate control, in which the transmission power in each radio frame is fixed while the data rate is adjusted by using different modulation and/or coding rates e.g. changes modulation schemes from QPSK (Quadrature Phase-Shift Keying) when the received SNR is poor and changes to 16QAM (16-State Quadrature Amplitude Modulation) when the received SNR becomes excellent.

An example of a system with power control (similar to typical CDMA systems), and rate control (similar to HSDPA/LTE systems) are illustrated in Figure 1.3.

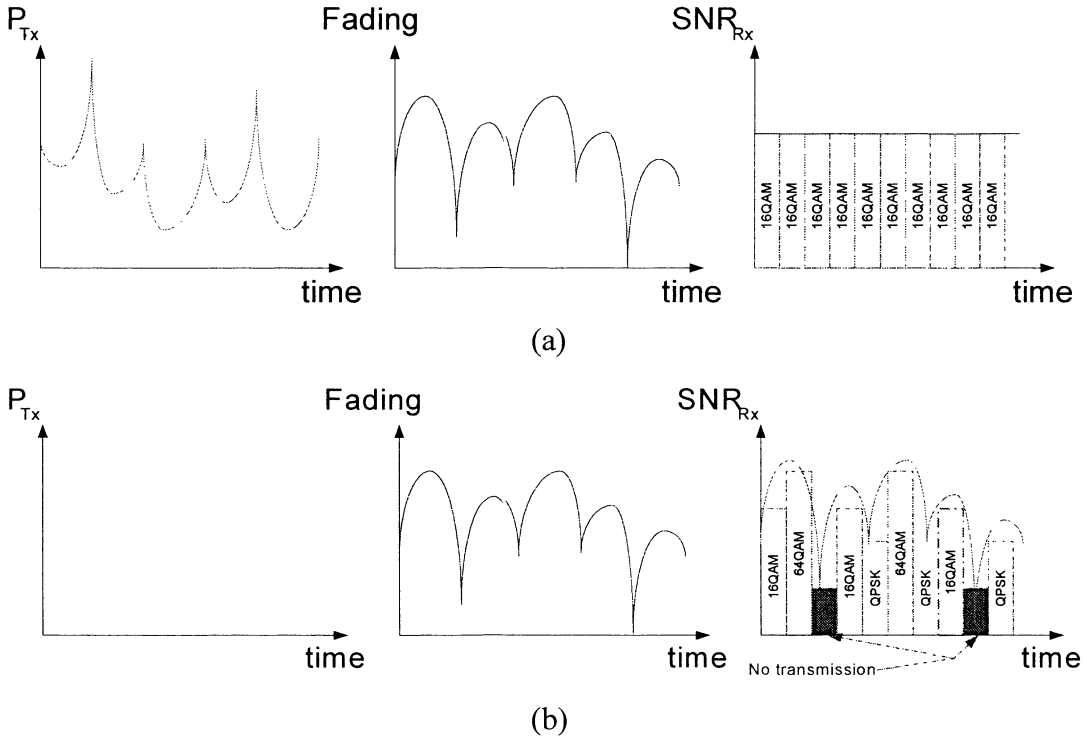


Figure 1.4 (a) Power control (b) Rate control

Figure 1.4 (a) shows that the transmission power in a power-controlled system is inversely proportional to the received SNR in order to maintain the received signal quality at a given target and thus the achievable data rate is approximately the same. On the other hand, Figure 1.4 (b) shows that the transmission power in a rate-controlled system is fixed, and the data rate is varying based on the received SNR. The power controlled transmission method is more suitable for traditional circuit-switched services where a constant data rate needs to be guaranteed while the rate controlled method is more efficient for packet-based services where highest data rate is preferred.

1.2.2 Multiuser Diversity and Opportunistic Scheduling

Diversity in telecommunication context refers to a transmission scheme in which information is sent over multiple channels so that if the data transmitted through one of the channels become corrupted, a large portion of information can successfully reach the destination. This concept is important in wireless communications as the radio channel is random in nature and thus there is always a

chance that one or more channels may become substantially poor at any point in time. Examples of classic diversity techniques in wireless networks are *frequency diversity* where data is transmitted over multiple frequency bands, *time diversity* where data transmission takes place on a number of time slots, and *antenna diversity* where multiple antennas are employed at the transmitter or receiver or both.

In wireless communications systems where multiple users are active, the quality of radio channel from each user to the base station is typically nonidentical to other users' link quality. The quality of the radio link is time-varying due to channel fading. It has been shown in the literature [8] that allowing users who have the highest received SNR among all users to have access to the resource maximises the achievable spectrum capacity, and this is referred to as *multiuser diversity* (*MuD*).

The research also discovered that the expected value of the highest received SNR among all the users consequently increases with a increase in the number of users [9]. The gain achieved by allowing users having their best received SNR to access the resource when the number of users is increased is called *multiuser diversity gain*. This gain makes multiuser diversity more attractive than the classic diversity techniques as multiuser diversity exploits the randomness in fading radio channel to improve the system capacity while the traditional diversity techniques were aimed to eliminate the effect of variation in channel fading which meant any gain from radio fading variation was discarded. The scheduling mechanism where users having the highest received SNR are allocated radio resource is referred to as *opportunistic scheduling* (*OpS*).

Figure 1.5 shows the received SNR of three users in a multiuser system employing an opportunistic scheduling where a user having the best received SNR, γ, in each scheduling slot is selected to transmit data.

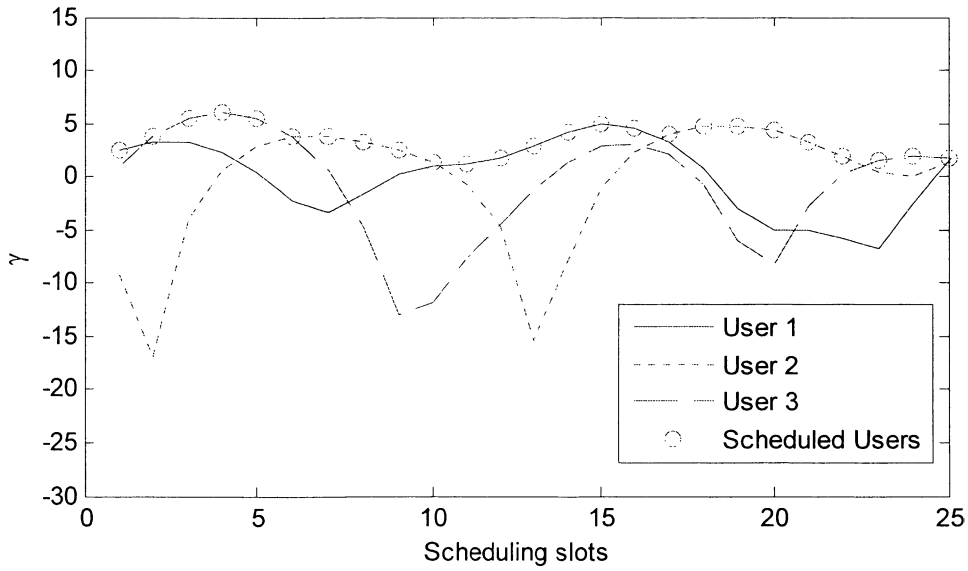


Figure 1.5 Multiuser system with opportunistic scheduling.

In Figure 1.5, User 1 is allowed to gain an access to the resource in scheduling slot 1. Then, User 3 is selected to gain an access to the resource from slots 2 to 6. After that User 2 is selected followed by User 1 as depicted in the diagram.

OpS is an important idea to achieve the multiuser diversity gain. However, the realisation of multiuser diversity gain comes with a requirement for considerable feedback overhead as the system would need to know the received SNR of all users so that users having the highest received SNR can be selected. This feedback requirement obviously demands substantial radio resource to send feedback and would make multiuser diversity gain less attractive to be implemented in practice. One of the research goals of this thesis is to develop new techniques to achieve multiuser diversity gain while minimising feedback resources required.

1.2.3 Channel State Information

To adaptively adjust the system parameters to suit the radio conditions, the transmitter needs the knowledge of radio channel conditions, which is referred to as channel state information (CSI), also known as channel quality indicator (CQI). CSI message contain information regarding the instantaneous radio channel conditions

such as received SNR or the supportable MCS mode of a user. In this thesis, the term CSI will be used throughout.

In frequency division duplexing (FDD) networks where uplink and downlink communications take place in different frequency bands, the requirement for CSI at the base station imposes the need for the users to take measurements and inform the base station of the instantaneous channels conditions so that the base station can adjust the system parameters accordingly to maximise the network performance. On one hand, inadequate CSI feedback would not allow the base station to effectively adjust the system parameters to suit the variation in channel conditions. On the other hand, excessive amount of CSI feedback might introduce substantial complexity to the system. More specifically, non-optimised CSI mechanism would increase the system complexity as all the users would have to calculate and send CSI messages on each scheduling time slot. Accordingly, the scheduler i.e. base station would have to unnecessarily process all CSI messages which results in high computational overhead of the system. Moreover, the amount of resource required for CSI feedback may exceed the amount of data to be sent. Hence, it is very important that the feedback is optimally generated and transmitted.

1.2.4 Hybrid Automatic Repeat-Request

Automatic Repeat-Request (ARQ) is an error-control protocol that is designed to provide reliable data transmission. In a system where ARQ is implemented, receivers will be required to send ACK message when the data is successfully received.

Hybrid Automatic Repeat-reQuest (HARQ) is a mechanism designed to detect, and correct errors that may appear over a communication link [10]. It makes uses of a combination of ARQ and forward error correction (FEC[1]) techniques. LTE supports two types of HARQ namely Chase combining (CC) and incremental redundancy

[1] FEC is an error-correction mechanism by which redundant messages are transmitted together with the data such that errors can be detected and corrected at the receivers.

(IR). In LTE, the information related to HARQ is contained in the scheduling grant message which is sent every downlink or uplink grant [11].

1.3 Problem Statements and Research Objectives

It has been shown in the literature that a system employing adaptive modulation and coding scheme (MCS) can achieve the capacity relatively closer to the Shannon's limit compared with those systems with constant transmission mode. The adaptive techniques require accurate channel knowledge at the base station so that it can accordingly select the optimum MCS. Improper link adaptation and CSI feedback design will result in excessive feedback requirements and this factor made OFDMA link adaptation not attractive in the past. Based on the aforementioned limitations, the following research questions are addressed:

- Given the importance of LAT in wireless systems, is it possible to design feedback reduction techniques that have low complexity and at the same time achieve a similar performance to full CSI [2] is available at the base station?

- Given the need to minimise feedback wireless systems, is it possible to develop link adaptation techniques that require fewer CSI feedback bits and at the same time achieve a similar performance as when full CSI is available at the base station?

- Provided that the number of feedback bits per each CSI message is limited to one bit, can the performance in terms of achievable sum-rate be further improved when compared with the existing algorithms?

- Provided that more feedback thresholds are introduced, can a mathematical expression of such feedback technique be formulated and if so how many thresholds are sufficient to obtain the full-CSI performance?

[2] Hereafter the theoretical achievable capacity provided that the transmitter has the perfect knowledge of CSI at the receiver is referred to as full-CSI.

- Provided that the best-M scheme when CSI's of best M subchannels out of all N subchannels are reported causes significant capacity loss when the number of users is small, is it possible to improve the algorithm so that it works well when the number of users is small?

- Provided that the feedback load increases with an increase of the number of users, how is it possible to develop a new media access control technique that can limit the feedback resources required regardless of the number of users?

1.4 Outline and Contributions

This thesis is organised as follows:

Chapter 1: Introduction

This chapter provides an introduction to the thesis. It contains brief information in cellular systems, LTE technology and a short description of opportunistic scheduling, link adaptation techniques in LTE which form part of the main content of this thesis. Furthermore, the problem statement and the aims of this work as well as research contributions are provided.

Chapter 2: Background

This chapter consists of background knowledge of LTE network architecture, OFDMA technology, adaptive modulation techniques, and system modelling. Basic scheduling algorithms are also briefly explained in this chapter.

Chapter 3: Link Adaptation with Reduced Feedback

This chapter contains an overview on feedback reduction techniques for multiuser system where link adaptation techniques are used. The implementation of feedback reduction techniques that contribute to the fundamental of feedback reduction techniques considered in this thesis is discussed in details. Then, a CSI reporting

technique in which the feedback is limited to two bits per CSI command for multiuser systems with proportional fair (PF) scheduling is proposed.

The work in this chapter was published and presented in:

[P1] R. Patachaianand and K. Sandrasegaran, "Proportional fair scheduling with reduced feedback", *IET Letter Vol 45*, *Issue 9*, April 2009.

[P2] R. Patachaianand and K. Sandrasegaran, "Opportunistic Feedback for Proportional Fair Scheduling", *Int. Conf. on Commun & Sig. Proc. ICCSP 2011*, Kerala, India, Feb 2011.

Chapter 4: Link Adaptation with Multiple Feedback Thresholds

In this chapter, the achievable sum rate of multiuser system with multiple feedback thresholds in a single carrier is formulated. This concept of multiple feedback thresholds is extended to a multiuser system with one-bit feedback.

The work in this chapter was presented in:

[P3] R. Patachaianand and K. Sandrasegaran, "Opportunistic Downlink Sum-Rate with Multiple Feedback Thresholds", *Int. Conf. on Compu. Appl. & Ind. Elect., ICCAIE 2010,* Dec 2010, Kuala Lumpur Malaysia.

[P4] R. Patachaianand and K. Sandrasegaran, "Link Adaptation with Multiple Feedback Thresholds", accepted for publication in IEEE, PIMRC, 201, 9-12 September 2012, Sydney Australia.

[P5] R. Patachaianand and K. Sandrasegaran, "Multiple Thresholds for One-Bit Feedback", in preparation.

Chapter 5: Automatic Feedback Threshold Setting and Hybrid Scheduling Techniques

The first part of the chapter presents a new technique to self-adjust the feedback threshold to suit the number of users. In the second part, a hybrid scheduling algorithm that behaves like PF scheduling policy when a CSI message is received, and resorts to the first-in-first-out (FIFO) policy in case of feedback outage is proposed.

The work in this chapter was presented in:

[P6] R. Patachaianand and K. Sandrasegaran, "Joint Delay-Aware Opportunistic Scheduling Algorithm with Reduced Feedback to Exploit Multiuser Diversity", *Int. Conf. on Compu. Appl. & Ind. Elect., ICCAIE 2010,* Dec 2010, Kuala Lumpur Malaysia.

[P7] R. Patachaianand and K. Sandrasegaran, "Hybrid scheduling technique for multiuser systems", *in Preparation.*

[P8] R. Patachaianand and K. Sandrasegaran, "Automatic Feedback Threshold Setting for Multiuser System with Opportunistic Scheduling", *in Preparation.*

[P9] R. Patachaianand and K. Sandrasegaran, "Hybrid Scheduling Technique to Reduce Scheduling Delay in Heavily Loaded Networks", *in Preparation.*

Chapter 6: Adaptive Feedback for OFDMA Systems

In this chapter, an adaptive feedback scheme that can self-adjust the bit-mask of feedback subchannels to suit the number of users in the coverage area is proposed.

The work in this chapter was presented and published in:

[P10] K. Sandrasegaran, R. Patachaianand, F. M. Madani, "Adaptive Feedback Algorithm for OFDMA Systems", *Proc. of IEEE Wir. Commu. & Netw. Conf, WCNC 2011*, March 2011, Quintana Roo, Mexico.

[P11] R. Patachaianand and K. Sandrasegaran, "Comparison of Feedback Reduction Techniques for OFDMA Systems", *in Preparation.*

Chapter 7: Opportunistic Feedback with Contention-based Channel

In this chapter, a new opportunistic feedback protocol for OFDMA systems with mixed realtime (RT) and non-realtime (NRT) traffic is proposed and evaluated.

The work in this chapter was presented and published in:

[P12] R. Patachaianand, K.Sandrasegaran, H. A. M. Ramli, and R. Basukala "Opportunistic Contention-Based Feedback Protocol for Downlink OFDMA with Mixed Traffic", *Springer Lecture Notes in Computer Science*, Volume 5879, Dec 2009.

[P13] K. Sandrasegaran, R. Patachaianand, F. M. Madani and C.C. Lin, "Analysis of Opportunistic Contention-Based Feedback Protocol for Downlink OFDMA", *Australasia Telecom. Network & Appl. Conf., ATNAC 2010*, Oct 2010, Auckland, New Zealand.

Chapter 8: Conclusions and future works

This chapter concludes this thesis and discusses some future work.

Chapter 2

Backgrounds

2.1 LTE Architecture

In the evolution from 1G to 3G and then finally to LTE, radio resource management (RRM) functions/mechanisms have moved from highly centralised (in 1G) to semi-distributed (in 3G) and then to totally distributed (in LTE) at the edges of the mobile network. In 1G mobile networks, RRM was centralized in the MSC (Mobile Switching Centre) resulting in a large amount of signaling in order to perform RRM functions. In 3G networks, RRM is semi-distributed in the Radio Network Controller (RNC) of the UTRAN. The RNC performs RRM functions for all the base stations (Node B) that are under its control. In LTE, 3GPP has standardised "evolved Node B" (eNB) [11] to replace Node B, and eNB will takeover all RRM functions from RNC and Node B.

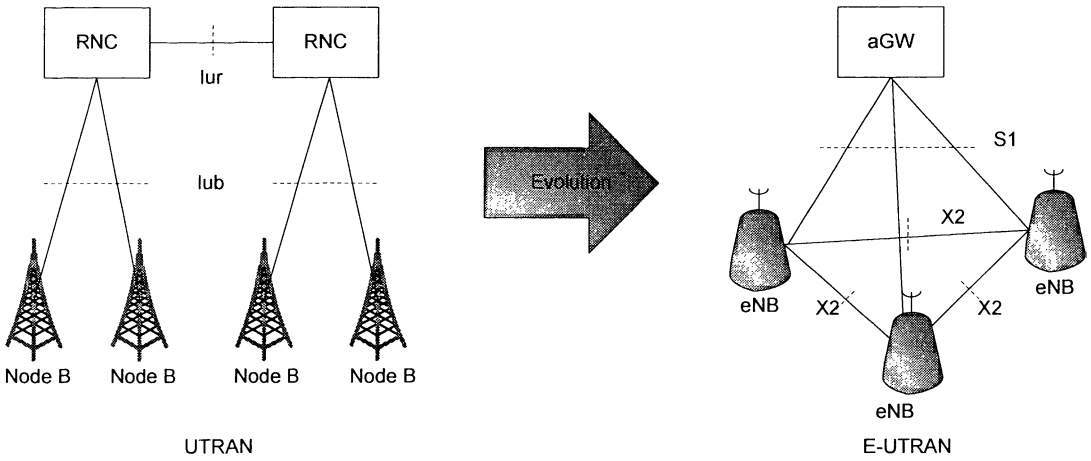


Figure 2.1 UTRAN and E-UTRAN architecture

The evolved UTRAN architecture is depicted in Figure 2.1, where eNB and aGW stand for eNode B and access gateway, respectively.

RRM functions to be performed at eNB are expected to be faster than in RNC because of the quicker exchange of measurements and signalling between the users and the radio access network. The fact that eNB is the entity that performs all RRM functionalities makes it feasible for advanced RRM techniques to be implemented because the signal does not have to travel a long distance as in traditional centralised or semi-distributed networks. However, the highly distributed nature of RRM is expected to create other problems, for example, high volume of signaling exchanges required for RRM. The performance of distributed RRM in LTE needs to be investigated carefully so that its benefits are realised. The investigation of the performance of RRM in eNB, particularly link adaptation (LAT) with efficient utilisation of signaling overhead is one of the major goals of this thesis.

2.2 LTE Physical Layer - OFDMA

Downlink communication in LTE is based on OFDMA technology where the radio resource is divided in time and frequency. The LTE downlink radio resource is referred to as resource block (RB)[3]. Each RB contains 12 subcarriers each of which has a subcarrier spacing of 15 kHz (total 180 kHz per RB). The LTE OFDMA sub-frame duration corresponds to 1 ms and each sub-frame is further divided into two equal 0.5 ms slots (equivalent to 7 OFDM symbols). The users can be allocated a RB in each sub-frame dynamically over time and frequency based on their CSI report and link adaptation criteria. The LTE OFDMA physical layer structure is illustrated in Figure 2.2.

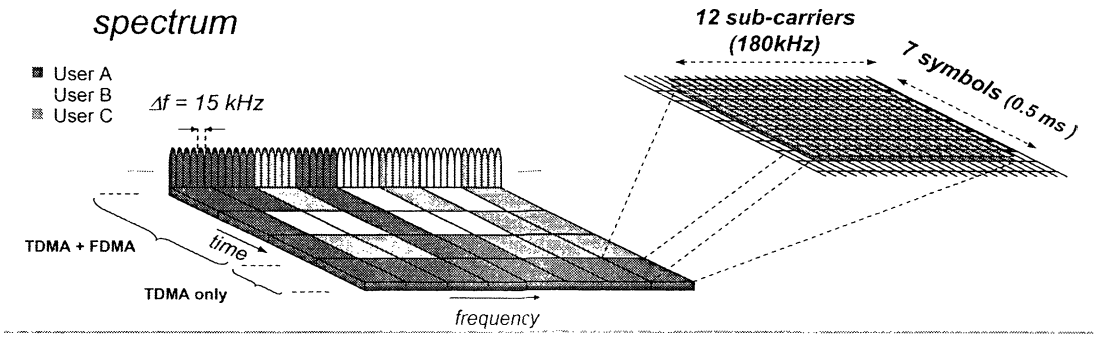


Figure 2.2 LTE OFDMA physical layer structure [4]

[3] Also referred to as physical resource block (PRB). These two terms are used interchangeably in this thesis.

[4] Acknowledged to Ericsson Research Tutorial documents

Figure 2.2 shows the LTE physical resource structure with dynamic resource allocation to three users (User A, User B, and User C). Each user may be allocated a RB in each 1 ms sub-frame where different MCS may change from one RB to another RB based on user channel condition. Different modulation and coding scheme (MCS) can dynamically change from one RB to another RB. More detail on adaptive MCS is discussed in the next section.

2.3 Adaptive Modulation and Coding Scheme

Adaptive modulation and coding scheme (MCS) can maximise the achievable spectral efficiency over time-varying wireless channels [12]. The basic idea of the adaptive MCS is the optimal MCS mode is selected at the transmitter (base station in downlink communications) to suit the receiver's instantaneous SNR level. To implement this, the receiver (mobile users) needs to estimate its SNR level and informs to the transmitter the current received SNR. The transmitter then uses this SNR information to optimally adapt the MCS mode to maximise the system capacity. It has been shown that multiple-QAM (MQAM) is efficient and can be implemented in practical systems [13][14]. Signal constellation degrees of 4, 16, and 64 QAM are given below:

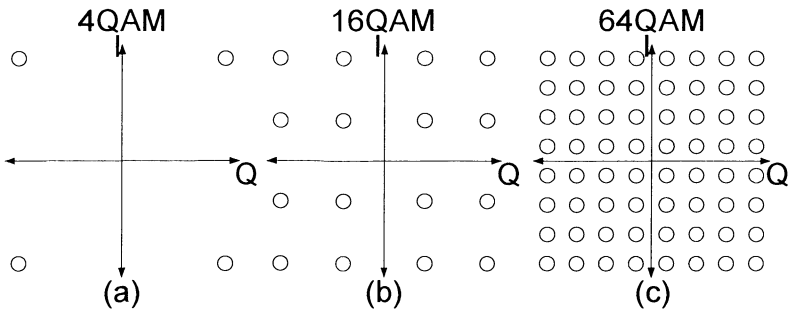


Figure 2.3 Different M-QAM

In Figure 2.3, one 4-QAM symbol carries two information bits while one 16-QAM symbol contains four information bits. This effectively means transmitting in 16-

QAM mode is two-times faster than 4-QAM modes. However, 16-QAM transmission is more sensitive to error than 4-QAM as illustrated below:

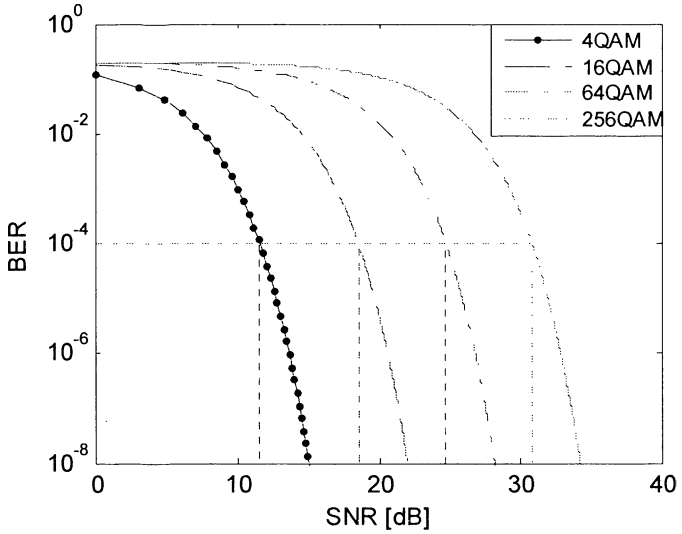


Figure 2.4 BER performance of uncoded MQAM

Figure 2.4 shows the bit error rate (BER) versus received SNR of different uncoded MQAM. It can clearly be seen that higher-order QAM requires higher SNR than a lower-order one to maintain the same BER. For example, at BER = 10^{-4}, 4QAM requires SNR = 12 dB while 16QAM requires 19 dB: 7 dB difference. The idea of MQAM has been known for decades, but could not be implemented in practical non-adaptive systems where modulation schemes are fixed because they require significantly high SNR to maintain BER at an acceptable value as illustrated below:

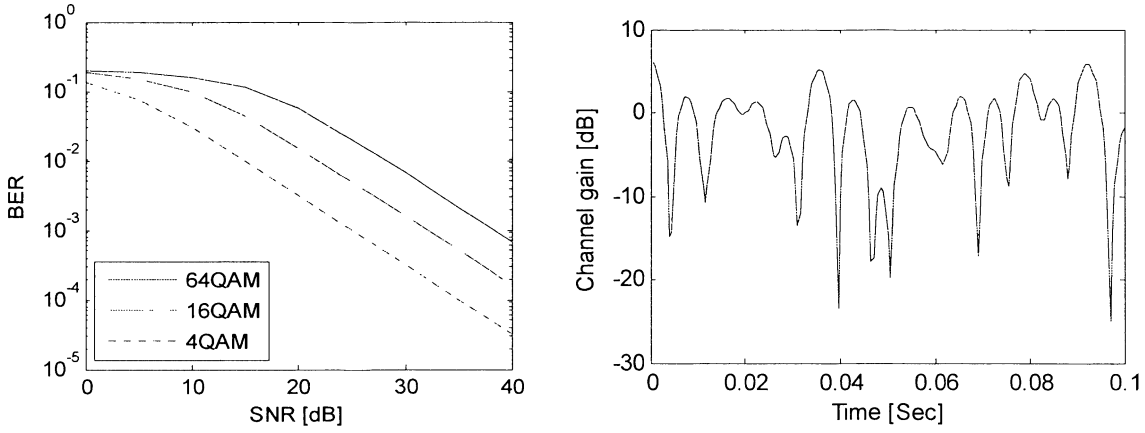


Figure 2.5 (a) BER performance of MQAM in Rayleigh fading channel (b) Typical Rayleigh channel

Figure 2.5 (a) indicates that, for 16QAM, a SNR of 40 dB is corresponding to a BER of approximately $3 \cdot 10^{-4}$. This implies that 16QAM requires more than 40 dB in received SNR level (not shown in Figure 2.5) to fulfill a requirement for BER = 10^{-4}. The fact that a higher order QAM requires higher SNR level to maintain the same BER as a lower order QAM in Rayleigh fading channel is due to a nature of radio channel. In particular, the fading can cause a substantial variation in received SNR, up to 30 dB difference from a peak SNR level and a lowest SNR level in a typical Rayleigh fading channel. Data transmission that takes place during the lowest SNR level is most likely to be erroneous and the error that occurs would dominate the average BER. To solve this, the system must be designed for the worst case scenario channel condition to guarantee a low BER throughout the entire duration of the communications. However, this design principle is very inefficient. In practice, it is difficult to design a system to cope with a variation of 30dB received SNR level while maintaining a satisfactory data rate due to several limitations such as transmit power limit regulation, excessive BER etc. Research work has shown that, to improve the overall capacity, a transmission when the received SNR is lower than a certain minimum level should be disabled and the transmission should only take place when the received SNR is sufficiently high [14]. Figure 2.6 shows the performance of adaptive modulation with 4QAM (SNR > 10 dB), 16QAM (SNR > 17 dB), and 64QAM (SNR >23 dB). The transmission is turned off if the SNR is less than or equal to 10 dB.

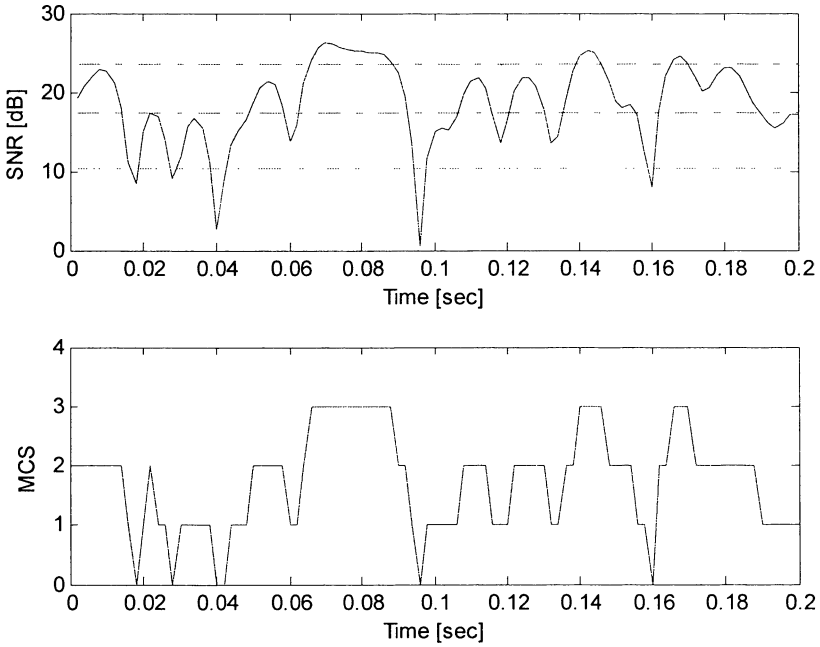


Figure 2.6 SNR level and MCS states mapping

Figure 2.6 shows a snapshot of received SNR and the adaptive MCS of a single-user system where the MCS 0, MCS 1, MCS 2, and MCS 3 correspond to no-transmission, 4QAM, 16QAM, and 64QAM, respectively. It can been seen from Figure 2.6 that adaptive modulation can allocate more bits per symbol using high-order QAM when the received SNR is high (such as at t = 0.07), and transmission is disabled when the channel quality is too low to maintain low-BER communications (such as t=0.04). This adaptive concept enables the system to achieve a high spectral efficiency when the average SNR of the system improves while the BER can be maintained to a target as shown in Figure 2.7 below. Figure 2.6 and 2.7 show the performance of a system with adaptive modulation scheme implemented.

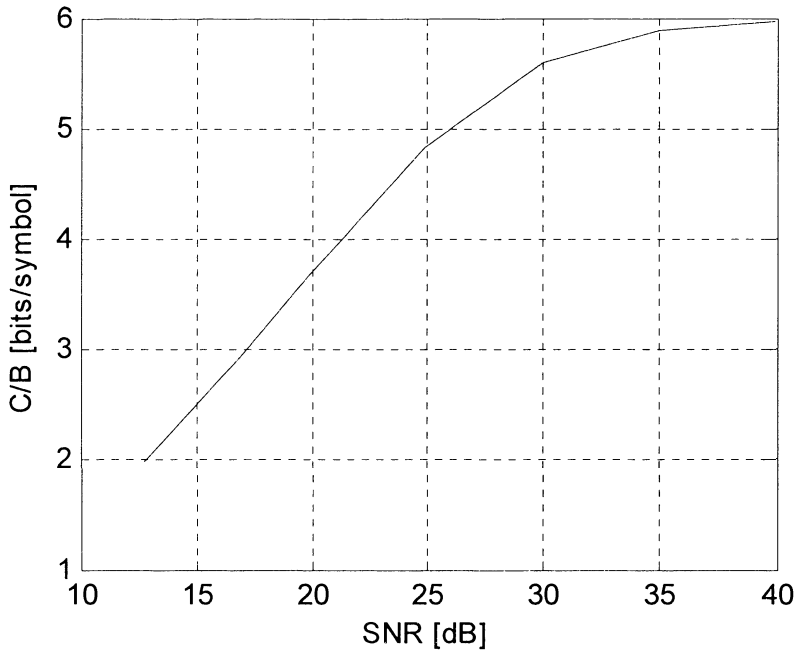


Figure 2.7 Spectral efficiency vs average SNR

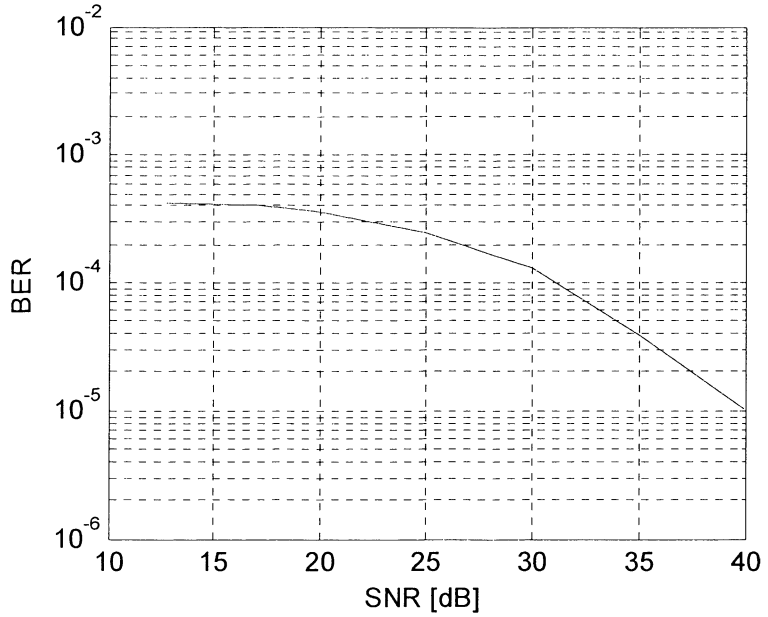


Figure 2.8 BER performance vs average SNR

2.4 Maximum Throughput and Proportional Fair Scheduling

Efficient scheduling algorithm can achieve multi-user diversity [9]. There are two common scheduling algorithms, namely, maximum throughput (MT) scheduling

and proportional fair (PF) scheduling [4]. MT scheduling maximises system sum-rate by allocating all resources to a user who has the highest received SNR in each scheduling slot. However, MT algorithm causes unfairness in a situation where mobile stations (MSs) are uniformly distributed over an area because users located closer to the base-station (BS) would have a higher chance to be scheduled than users located far away from the BS On the other hand, PF algorithm can maintain fairness among all MSs regardless of their location but it has a relatively smaller sum-rate compared with MT.

In MT and PF algorithms, all the available resources would be allocated to a user who has the highest scheduling metric in a scheduling time slot. The scheduling metrics, $z_i(t)$, for MT and PF are calculated as follows [15][16]:

$$z_i(t) = \begin{cases} \gamma_i(t) & ; MT\ scheduling \\ \gamma_i(t) / \overline{\gamma}_i(t) & ; PF\ scheduling \end{cases} \tag{2.1}$$

where $\gamma_i(t)$ and $\overline{\gamma}_i(t)$ denote received signal-to-noise ratio (SNR) and average SNR of i^{th} user at t^{th} scheduling time slot (moving average), respectively. The scheduler in a system employing MT or PF needs to know the received SNR of the users so that the scheduling can be performed efficiently. In a downlink system, MSs estimate the received SNR and delivered to the corresponding BS by means of channel station information (CSI) via a feedback channel. CSI information sent from every MS at every scheduling interval would cause excessive overhead to the system because only one user having the highest metric will be selected, and CSI from all other users would be discarded.

2.5 CSI Feedback for Link Adaptation

In order to realise the adaptive modulation, the transmitter requires the channel state information (CSI) feedback from the receiver so that it can select an optimal MCS mode to suit current channel quality. A diagram showing the concept of adaptive modulation with feedback is given below:

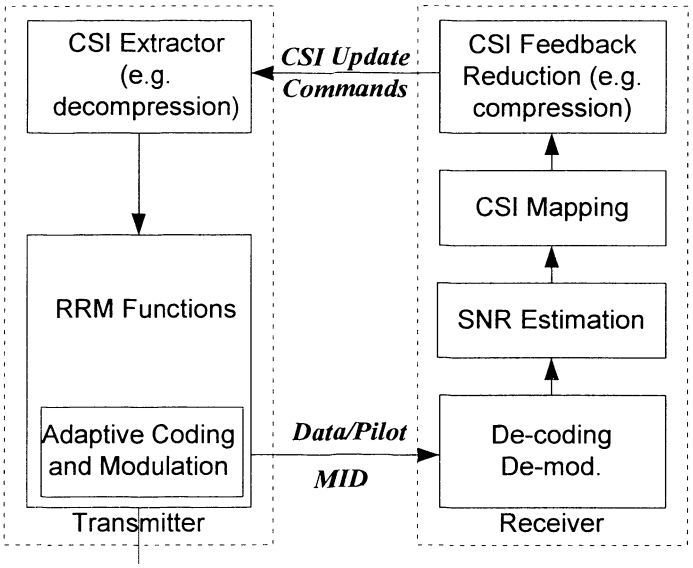


Figure 2.9 Typical adaptive modulation system with feedback

In Figure 2.9, the transmitter sends data signal together with pilot signal to the receiver. It is important for the transmitter to send MCS index (MID) to indicate the MCS that is currently in use so that the receiver can appropriately demodulate the received signal. Once the receiver receives the signal, it estimates received SNR from the pilot signal, and passes the estimated SNR to SNR-to-CSI mapping to map between SNR and corresponding CSI. As the radio resource blocks in OFDMA system span over time and frequency domains, a CSI message would contain more information than in typical single carrier systems and therefore needs to be carefully optimised. Figure 2.10 shows the typical frequency selective Rayleigh fading channel.

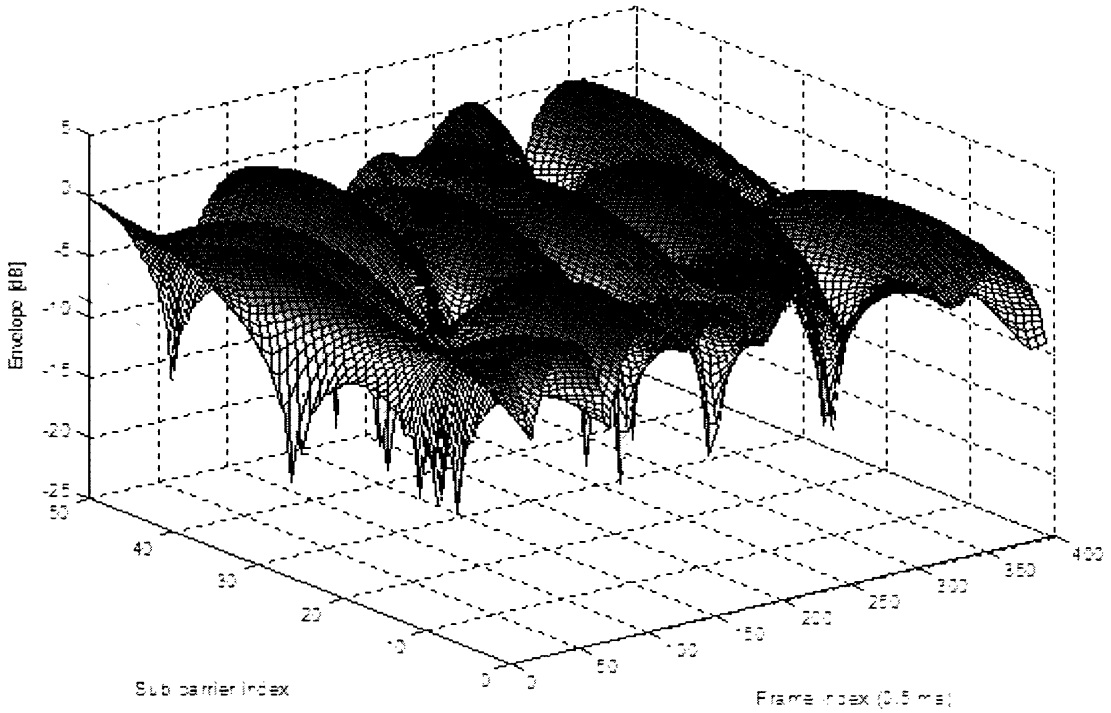


Figure 2.10 Frequency selective fading

From Figure 2.10, it can be observed that the channel quality varies in both time and frequency domains (resource blocks in OFDMA cases).

In downlink communication (base station to mobiles), mobile stations can perform accurate channel estimation based on the pilot signal received from the base station. In FDD systems where dynamic LAT is implemented, channel knowledge is sent from mobile stations to the base station by means of channel state information (CSI). The feedback load required for delivering CSI is critical for OFDMA systems such as LTE because there are many degrees of freedom to be adapted such as resource block sizes, and modulation schemes. Consider an OFDMA system with K users each of which is equipped with a single antenna, L CSI levels, and N subchannels. In the full CSI scenario, the amount of feedback bits to be sent in each scheduling slot is

$$FB = K \cdot N \cdot B \qquad (2.3)$$

where $B = ceil(log_2(L))$ is the number of bits to represent L CSI states (typically L = 32 i.e. $\log_2(L) = {}_B = 5$) and ceil(x) is the minimum integer number greater than or equal to x. Let assume that $N = 25$, and $L = 32$ the amount of feedback to be sent per

user in each scheduling slot is 25·5 = 125 bits per scheduling slot per user. The scheduling in LTE is performed on 1 ms basis. This results in a feedback rate requirement of 125 kbps in the uplink direction per user. From this simple example, it can be observed that an inefficient CSI feedback mechanism will result in excessive feedback overhead and thus consumes a large portion of the system radio resource.

The work in this thesis focuses on optimising different parts of the feedback resources as summarised in Table 2.1 and Figure 2.11.

Table 2.1 Summary of contributions regarding to feedback reduction aspects

	Number of feedback users, K_{FB}.	**Bits to represent number of resource blocks, N.**	**Bits to represent CSI levels, $log_2(B)$.**
Chapter 3	Reduction of number of users sending feedback by optimally selected feedback threshold.	N/A	Reduce number of bits required to represent CSI from 4 bits to 2 bits.
Chapter 4	Further reduce number of users sending feedback by optimally selected multiple feedback thresholds in comparison with single feedback case.	N/A	Further utilise one bit CSI to improve the achievable sum rate.
Chapter 5	A feedback reduction technique that can self-adjust the feedback threshold to reduce the number of users sending feedback.	N/A	N/A
Chapter 6	N/A	A feedback reduction technique that can self-adjust CSI bitmap of resource blocks to suit the number of users.	Feedback reduction technique to eliminate the need for CSI to represent SNR/MCS levels by using HARQ messages.
Chapter 7	Technique to reduce the number of users sending feedback in OFDMA systems.	Optimal selection of number of resource blocks to report CSI is determined.	N/A

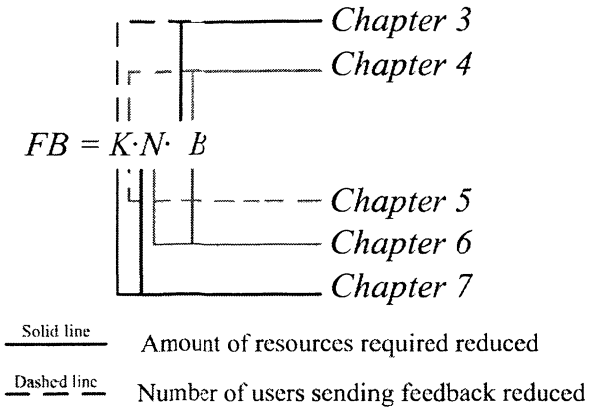


Figure 2.11 Summary of focus of this thesis

2.6 Radio Resource Management Functions

In a downlink transmission, majority of the RRM functions take place at the base station (eNB in LTE). A block diagram showing the interfaces between different RRM entities in performing dynamic link adaptation and resource allocation in OFDMA based systems is shown in Figure 2.12 [17].

In Figure 2.12, the base station receives CSI and ACK/NACK messages from the users. The CSI messages are then passed to the link adaptation (LA) entity. The LA unit consists of two main functions, namely, outer-loop LA and inner-loop LA. The inner loop LA is responsible to determine the supportable modulation and coding scheme based on received CSI. The outer-loop LA is responsible for adjusting the function to map between estimated SNR to CSI in order to maintain the average BER under a decided value in a similar manner as outer-loop power control in WCDMA-UMTS [18].

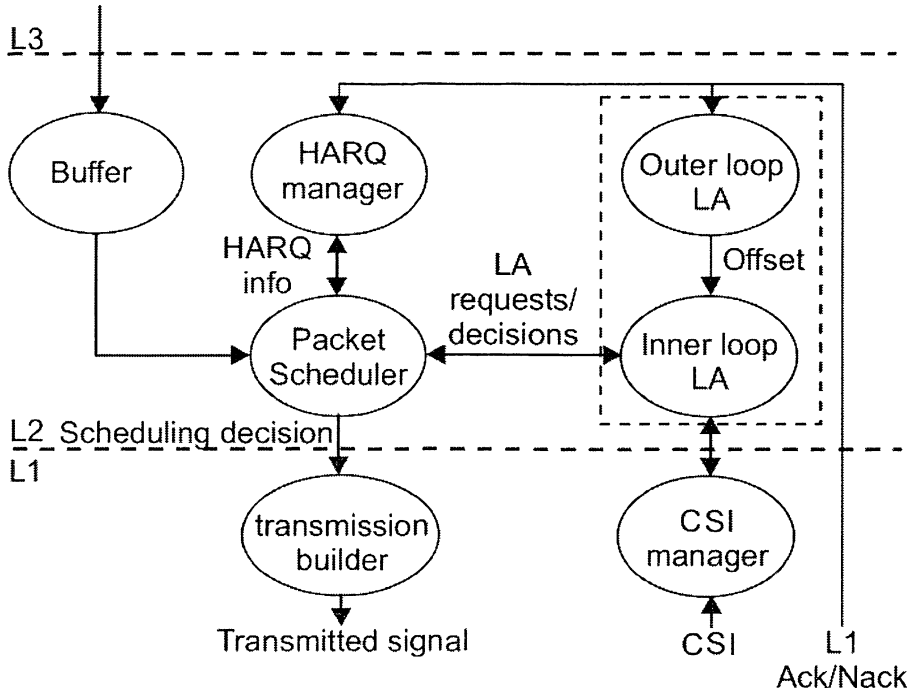


Figure 2.12 Interfaces between radio resource management functions [17]

L1 in Figure 2.12 represents functions that are related to physical layer such as transmission builder signal transmission, and CSI manager.

2.7 System Modeling

The general system model used in this thesis is described in this section. More specific details are presented in the corresponding chapter and will override the general system model described in Section 2.7.1 to 2.7.4.

2.7.1 Channel Model

The downlink data transmission of a single-cell system with K active users communicating with a base station (BS) employing a scheduling algorithm to select a user in each time slot is considered in this thesis.

The channel (i.e. subchannel in OFDMA system) is modelled as frequency-flat block-Rayleigh fading where the fading is assumed to be constant over a

transmission block. At any arbitrary time, the data transmission of the k^{th} user can be modeled as:

$$y_{lk} = h_k x + w_k \tag{2.15}$$

where y_k is the received signal, h_k is the fading gain $h_k \sim \mathcal{CN}(0,1)$, and w_k is the additive white Gaussian noise (AWGN) $x_k \sim \mathcal{CN}(0, \sigma)$. The normalised signal-to-noise ratio can then be computed by $\gamma_k = |h_k|^2/\sigma^2$. Given that the thermal noise power density $\sigma^2 = 1$, the probability density function (PDF) of γ_k follows:

$$f_{\gamma_k}(x) = \frac{1}{\bar{\gamma}_k} \cdot e^{-\frac{x}{\bar{\gamma}_k}}; \quad x \geq 0. \tag{2.16}$$

where $\bar{\gamma}_k$ is the average SNR of k^{th} user.

It is assumed that the fading between different users are independent. An example of Rayleigh fading channel was plotted in Figure 2.5.

2.7.2 Data Rate

The achievable spectral efficient is determined by the Shannon's capacity formula i.e.:

$$C_k(\gamma_k) = \log_2(1+\gamma_k) \quad \text{bps/Hz} \tag{2.17}$$

where γ_k is the signal-to-noise ratio (SNR[5]) of the k^{th} user.

The capacity in Equation 2.17 can be used as a measure of the maximum achievable transmission rate per bandwidth that can be supported by the base station when the transmission format is optimally selected to suit the scheduled user's received SNR in each scheduling slot. This method is an approximation over a more practical case where the transmission format is selected from a finite set of discrete

[5] SNR in this thesis is referred to as the ratio between the level of desired signal to undesired signals which may also referred to as signal to interference plus noise ratio.

modulation and coding schemes (MCS) for data transmission similar to the mapping method used in Table 2.2 below and Figure 2.13 to 2.15.

Table 2.2: SNR to MCS mapping table [19]

MCS mode	SNR (dB)	Data rate (kbps)
0	< -12.5	0
1	-12.5	28.4
2	-9.5	76.8
3	-8.5	102.6
4	-6.5	153.6
5	-5.7	204.8
6	-4	307.2
7	-1	614.4
8	1.3	921.6
9	3	1288.8
10	7.2	1843.2
11	9.5	2457.6

Data rate in Table 2.2 is from HDR system and has been achieved by a combination of QPSK, 8-PSK, and 16-QAM with 1/5, 1/3, and 2/3 turbo coding. The system bandwidth is configured at 5 MHz.

The relationship between MCS modes and SNR in Table 2.2 is plotted in Figure 2.13 below:

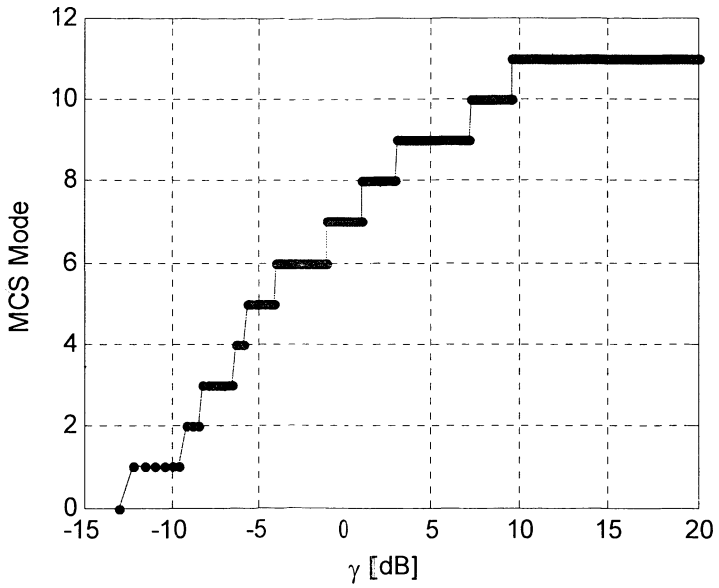


Figure 2.13 Relationship between SNR modes and MCS

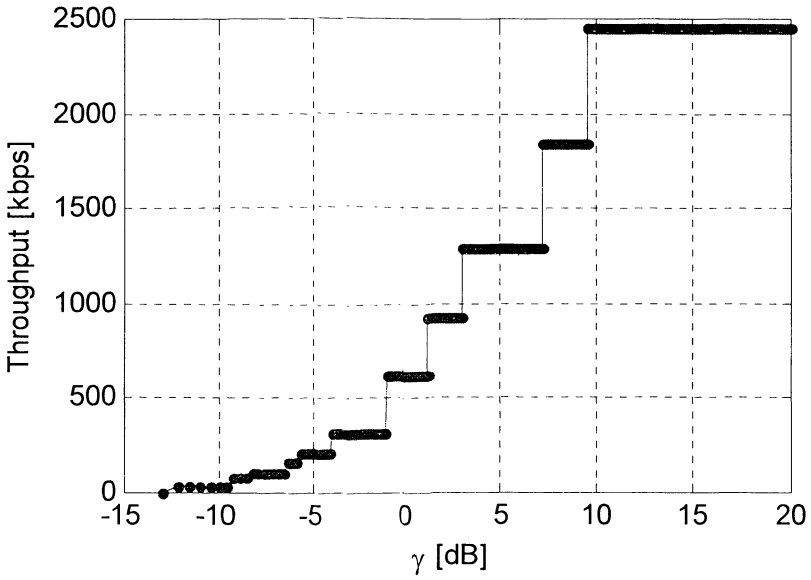


Figure 2.14 Relationship between SNR and supportable throughput

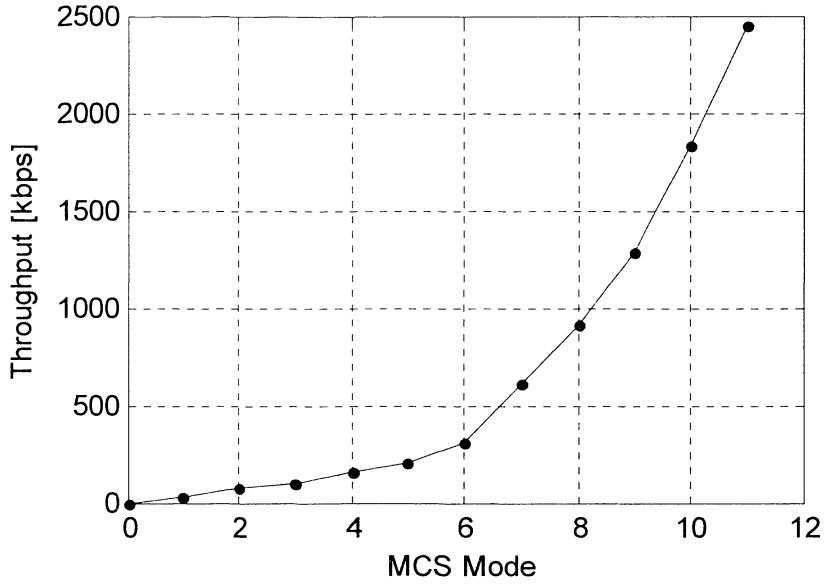


Figure 2.15 Relationship between MCS mode and supportable throughput

2.7.3 Scheduling Algorithms

Proportional fair scheduling is used in this thesis except in Chapter 4 and 8 where the MT scheduling is used. The scheduling is performed on a slot-by-slot basis i.e. at most one user may have access to each scheduling slot in each channel (or subchannel in OFDMA). At most one user may be scheduled in each scheduling slot. With PF scheduling policy, the scheduled user, $k^*(t)$, is the one having the maximum normalised capacity in scheduling slot t i.e.:

$$k^*(t) = \arg\max_{k=1\ldots K}\left\{\frac{r_k(t)}{R_k(t)}\right\} \tag{2.18}$$

where $r_k(t)$ is achievable instantaneous data rate at time t as shown in (2.17), and $R_k(t)$ is the average data rate of the k^{th} user at time t. The data rate in each scheduling slot is updated based on the following equation [9]:

$$R_k(t+1) = \begin{cases} \left(1-\frac{1}{t_c}\right)\cdot R_k(t) & k \neq k^*(t) \\ \left(1-\frac{1}{t_c}\right)\cdot R_k(t) + \frac{1}{t_c}\cdot r_k(t) & k = k^*(t) \end{cases} \tag{2.19}$$

where t_c is the window-averaging time constant (in this thesis, $t_c = 2000$ is used).

In the long term when t approaches a large value, the performance of PF scheduling approaches to allocating the resource to a user having the highest relative SNR (received SNR divided by average SNR) in each scheduling slot [15] i.e.:

$$k^*(t) = \arg\max_{k=1...K}\left\{\frac{r_k(t)}{R_k(t)}\right\} \approx \arg\max_{k=1...K}\left\{\frac{\gamma_k(t)}{\bar{\gamma}_k(t)}\right\} \tag{2.20}$$

Scheduling technique using the criteria in Equation 2.20 is defined in this thesis as Simplified PF. The achievable sum rate, C, is defined as the average spectral efficiency of the multiuser system employing opportunistic scheduling (i.e. PF scheduling), and can be computed as:

$$C = \int_0^\infty C_{k^*}(\gamma) f_{k^*}(\gamma) d\gamma \qquad \text{bps/Hz} \tag{2.21}$$

where C_{k^*} is the spectra efficiency as expressed in Equation 2.17 and $f_{k^*}(\gamma)$ is the probability density function of the highest relative SNR among all user K.

2.7.4 Feedback Scheme

Every mobile estimates its SNR in every scheduling interval but only some mobiles transmit CSI message to the base station through feedback channel. In particular, for system with PF scheduling policy, only mobile users whose relative SNR is greater than a feedback threshold are allowed to send CSI i.e. user k^{th} having:

$$\frac{\gamma_k(t)}{\bar{\gamma}_k(t)} \geq \gamma_{th} \tag{2.22}$$

where $\bar{\gamma}_k(t)$ is the average SNR of user k^{th} and γ_{th} is the feedback threshold. The term $\gamma_k(t)/\bar{\gamma}_k(t)$ is defined as relative received SNR (or in short 'relative SNR') of user k^{th}.

It is assumed that user always send user's identification (user ID) associated with CSI sent. As bits to represents user ID is the same for all feedback schemes and may be used for purposes other than CSI reporting, user ID bits are not considered in feedback load calculation in this thesis.

Feedback channels are assumed to be errorless and delay-free.

2.7.5 Network Load

Throughout this thesis, the network are assumed to be fully loaded i.e. the base station always has packets queued within buffers to be sent to each user. The definition of heavily loaded network is where the network is fully loaded and less than 10% of users can be served at each scheduling load.

2.8 Summary

In this chapter, background information related to the content of this thesis is overviewed. In particular, the LTE system architecture, link adaptation techniques and CSI feedback mechanisms are discussed. Furthermore, functions of RRM and details of system modelling are briefly described.

Chapter 3

Link Adaptation with Reduced Feedback

This chapter evaluates the performance of the feedback reduction techniques proposed in the literature for multiuser single carrier system. Furthermore, a new feedback reduction technique that aims to minimise the number of bits in each CSI message is proposed.

This chapter begins with a brief introduction to feedback reduction techniques in Section 3.1. In Section 3.2, feedback reduction techniques presented in the literature are overviewed. The effect of feedback thresholds on the sum rate performance is discussed in Section 3.3. In Section 3.4, modified versions of selective feedback and one-bit feedback techniques for proportional fair scheduling are introduced. New feedback reduction technique is proposed in Section 3.5. Simulation results showing the performance of the proposed feedback reduction technique are presented in Section 3.6. Section 3.7 summaries this chapter.

3.1 Introduction

In single carrier systems where opportunistic scheduling is used, it has been shown that allocating the radio resource to a single user whose received SNR is the best among all users in each scheduling slot achieves optimum sum-rate capacity [9]. To perform opportunistic scheduling, the scheduler at the base station requires knowledge of CSI from all users.

To generate CSI messages, each mobile user estimates the received SNR in each scheduling slot. This estimated SNR value is then mapped to an appropriate CSI value which may represent a quantised SNR value or MCS value depending on the system configuration. In this chapter, estimated SNR will be mapped to an appropriate MCS as per Table 2.1. There are 12 MCS modes available so that, for a

standard MCS-to-CSI mapping, one CSI command contains at least four bits ($\lceil \log_2(12) \rceil = 4$; where $\lceil x \rceil$ is an integer which is greater than or equal to x.

CSI information from all users in every scheduling slot represents a considerable feedback load because scheduling is normally performed on a fast basis (typically in order of 1 ms). This fast scheduling implies that the CSI will need to be sent in the same rate to ensure that an efficient scheduling can be performed. As at most one user may be scheduled in each scheduling slot, and hence most of the CSI would be wasted. In other words, majority of the users would not be scheduled to have access to the radio resource and thus it is not necessary for them to report CSI.

Research work for feedback reduction in single carrier systems can be classified into three categories as below:

- Full CSI: All users sending CSI which contains a real number to represent received SNR. This is typically used in mathematical analysis and as a performance benchmark.
- Threshold based feedback reduction scheme: Only users whose received SNR is above a predefined feedback threshold are allowed to send feedback. This scheme reduces the number of users sending feedback, while information contained in CSI may be several bits.
- Quantised feedback scheme: This scheme focus on minimising the amount of bits contained in CSI to represent estimated SNR or supportable MCS.

In this chapter, a new feedback reduction technique that is a combination of threshold based and quantised feedback schemes are the main focus.

3.2 Overview of Feedback Reduction Techniques

The previous work to achieve multiuser diversity gain using reduced feedback in the literature that contributes to the fundamentals of this chapter is described in the following subsections.

3.2.1 Gesbert's Selective Feedback

Gesbert et al [20] suggested that allowing users to report CSI only when their received SNR are higher than a predefined threshold, $\gamma_{th,}$ can reduce feedback load considerably while maintaining the achievable sum-rate near optimum. The selection criteria of Gesbert's selective feedback can be expressed as [20],

$$\begin{aligned} k^*(t) &= \arg\max_{k=1\ldots K}\{\gamma_k(t)\} \quad \textit{when at lest one user sends feedback} \\ k^*(t) &= \underset{k=1\ldots K}{rand}\{k\} \quad \textit{when none of users sends feedback} \end{aligned} \tag{3.1}$$

where rand$\{k\}$ is a uniformly random selection of variable k. $\gamma_k(t)$ is the received SNR.

At scheduling slot t, user k will send CSI feedback to the base station if and only if [20]:

$$\gamma_k(t) \geq \gamma_{th} \tag{3.2}$$

where γ_{th} is the feedback threshold.

The sum rate capacity of selective feedback can be expressed as [20]:

$$\begin{aligned} E(C_*) &= \log_2(e)\left(1-e^{-\gamma_{th}/\overline{\gamma}}\right)^{K-1} \\ &\left[e^{1/\overline{\gamma}}\left(E_1\left(\frac{1}{\overline{\gamma}}\right)-E_1\left(\frac{1+\gamma_{th}}{\overline{\gamma}}\right)\right)-e^{-\gamma_{th}/\overline{\gamma}}\ \ln(1+\gamma_{th})\right] \\ &+\log_2(e)\sum_{k=1}^{K}\binom{K}{k}\,k\left(1-e^{-\gamma_{th}/\overline{\gamma}}\right)^{K-k} \\ &\sum_{n=0}^{k-1}\binom{k-1}{n}(-1)^n\,\frac{e^{(-k+1+n)\gamma_{th}/\overline{\gamma}}}{n+1} \\ &\left[e^{-(n+1)\gamma_{th}/\overline{\gamma}}\ \ln(1+\gamma_{th})+e^{\frac{n+1}{\overline{\gamma}}}\ E_1\left(\frac{(n+1)(1+\gamma_{th})}{\overline{\gamma}}\right)\right] \end{aligned} \tag{3.3a}$$

where

$$
\begin{aligned}
p_{\gamma *}(\gamma) &= (P_\gamma(\gamma_{th}))^{K-1}\ p_\gamma(\gamma), \qquad \gamma \le \gamma_{th} \\
p_{\gamma *}(\gamma) &= \sum_{k=1}^{K} \binom{K}{k} (P_\gamma(\gamma_{th}))^{K-k}\ k\, p_\gamma(\gamma) \\
&\quad \times (P_\gamma(\gamma) - P_\gamma(\gamma_{th}))^{k-1}, \qquad \gamma > \gamma_{th},
\end{aligned} \tag{3.3b}
$$

and

$$E_1(x) = \int^{\infty} e^{-xt}/t\, dt \tag{3.3c}$$

The selection of feedback threshold plays an important role on the performance of multiuser systems with limited feedback. More specifically, inappropriately selected feedback threshold will result in the performance loss of the achievable system capacity.

3.2.2 One-bit feedback

Sanayei et al [21] proposed a feedback scheme by which each of the users whose received SNR are higher than a threshold is allowed to send one-bit feedback to the base station. The base station then selects a user to be scheduled randomly out of K_{FB} users sending one-bit CSI. Mathematical analysis in [21] showed that opportunistic scheduling in multiuser systems with the one-bit feedback scheme can achieve the same growth rate in terms of achievable capacity as full-CSI when the number of users increases. Somekh et al [22] obtained a mathematical expression for the capacity growth rate in multiuser system with one-bit feedback as in [21] using a different approach.

The closed-form expression of the sum-rate, C, is given by [22] :

$$C(\gamma_{th}, K, \bar{\gamma}) = Q(\gamma_{th}, K) \cdot \chi(\gamma_{th}, \bar{\gamma}) \tag{3.4}$$

where

$$Q(\gamma_{th}, K) = (1 - (1 - e^{-\gamma_{th}})^K) \tag{3.5a}$$

$$\chi(\gamma_{th}, \bar{\gamma}) = \log(1 + \gamma_{th}\bar{\gamma}) + e^{\left(\frac{1+\gamma_{th}\bar{\gamma}}{\bar{\gamma}}\right)} \cdot Ei\left(\frac{1+\gamma_{th}\bar{\gamma}}{\bar{\gamma}}\right) \tag{3.5b}$$

$Ei(x) = \int_x^\infty e^{-t}/t\,dt$ is the exponential integral function and *log(x)* is a natural logarithm of a variable *x*. $\bar{\gamma}(t)$ is the average SNR.

The optimal threshold, $\hat{\gamma}_{th}$, can be selected such that the following condition is satisfied:

$$\hat{\gamma}_{th}(K,\bar{\gamma}) = \arg\max_{\gamma_{th}} C(\gamma_{th}, K, \bar{\gamma}) \tag{3.6}$$

The one-bit feedback technique can further reduce the feedback load more effectively than Gesbert's technique. Nevertheless, the one-bit technique results in non-negligible capacity loss compared with the full-feedback as well as Gesbert's technique (the performance of these schemes are compared in Section 3.5).

3.2.3 Kim's Feedback Reduction

Kim et al [23] proposed a feedback load reduction technique that automatically induces select users to send CSI only when the probability to be scheduled is relatively high. The feedback threshold is dynamically updated in each scheduling slot as following [23]:

$$\gamma_{th} = \begin{cases} \gamma_{th} + \Delta\gamma_{th} & ;if\ k_F > k_T \\ \gamma_{th} - \dfrac{1-\phi}{\phi} \cdot \Delta\gamma_{th} & ;else \end{cases} \tag{3.7}$$

where

k_F is the number of users sending feedback

k_T is the target number of users sending feedback,

$\Delta\gamma$ is the increasing factor for feedback threshold;

and ϕ is a real number set to optimise the probability of users sending feedback.

The feedback threshold update policy by Kim et al requires the knowledge of the number of users sending feedback in each scheduling. This update policy implies that the base station needs to continuously inform to the users the feedback threshold which consequently introduces additional load to the system.

Furthermore, Kim's technique requires users to send CSI containing a full range of MCS mode to the base station which implies that the feedback does not reduce from the selective feedback as previously proposed in [20].

3.2.4 Floren's Feedback Reduction

Floren et al in [24] focused on optimising the feedback quantisation for a multiuser system with constant-rate transmission. It was suggested that the user SNR should be quantised and represented by a few bits. The sum rate of multiuser system with quantised CSI , C_{QMS},can be represented as:

$$C_{QMS} = \frac{1}{2} \cdot \sum \int_{Q_j} \log_2(1+\gamma) \cdot \frac{[F_\gamma(q_{i+1})]^K - [F_\gamma(q_i)]^{FK}}{K \cdot (F_\gamma(q_{i+1}) - F_\gamma(q_i))} \cdot f_\gamma(\gamma) d\gamma \tag{3.8}$$

where Q_i is the i^{th} quantisation region, $F_\gamma(x)$ is the cumulative distribution function of received SNR, f_γ is the probability density function of received SNR, K is the number of users, and q_i is the values of the SNR that limit the i^{th} quantization level.

However, the number of users sending feedback was not considered and thus the amount of feedback might become excessively large when the number of users grows. The simulation results in [24] did not clearly reflect the multiuser diversity gain achieved by the proposed quantisation technique.

3.2.5 Holter's Feedback Reduction

In [25], Holter et al presented a multiuser access scheme to minimise the average feedback load. Holter's feedback scheme allows the base station to probe each user with a threshold to check whether its SNR is greater than the threshold. If the SNR is below the threshold, the base station moves to the next user. The base station continues probing/checking for users' SNR until one of the following events occurs:

- Until it detects a user with SNR above the threshold, it will then select this user the next transmission time.

- Until all users have been probed and none of them has SNR that exceeds the threshold. In this case, a user having the highest SNR will be selected, or the base station may wait for a period longer than the channel coherence time and starts a new sequential search.

Although Holter's feedback reduction technique showed some performance improvement in terms of feedback load and spectral efficiently improvement, it requires significant probing period. In addition, the users might have to participate in numerous SNR checking during the probing.

3.3 Effect of Feedback Threshold

In this section, the effect of feedback threshold selection on the sum rate performance for the Gebert's selective feedback and one-bit feedback is discussed.

3.3.1 Effect of Feedback Threshold for Gesbert's Selective Feedback

Let us consider a multiuser system with $K = 10$ users and Gesbert's selective feedback technique is applied. Each user whose received SNR is above the feedback threshold would send a CSI message that contains B bits to represent all possible range of MCS. Gesbert's selective feedback technique will be referred to as B-bits selective feedback scheme hereafter. In this chapter, discrete MCS as shown in Table 2.1 is used in the simulation. There are 12 MCS levels i.e. CSI contains ($\lceil \log_2(12) \rceil = 4$ bits so that B = 4. Figure 3.1 shows the achievable net throughput of a multiuser system with 10 users as a function of feedback threshold. The average received SNR for all users is the same at 0 dB.

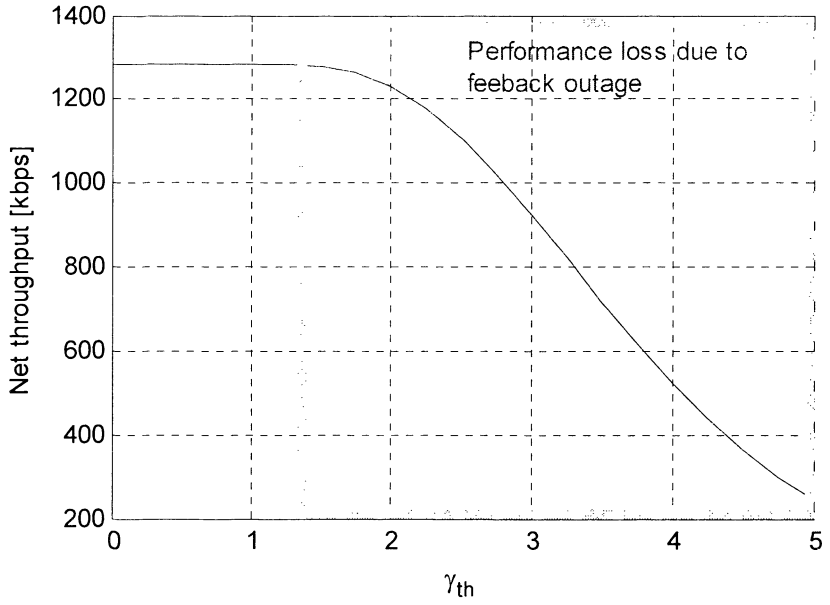


Figure 3.1 Performance of multiuser system with B-bits selective feedback

Figure 3.1 demonstrates that the achievable net throughput reduces with an increase of the value of the feedback threshold. This is due to the fact that with the B-bits selective feedback scheme, there is a possibility that all users may have their received SNR below the threshold simultaneously in a scheduling slot. As a result, none of users sends CSI feedback and the base station has to select a user to be scheduling in this particular scheduling slot without any knowledge of CSI. This situation is referred to as *feedback outage*. The feedback outage probability, P_{ot}, is a function of the number of user, K, and the value of feedback threshold, γ_{th}, as expressed in Equation 3.6 [20]:

$$P_{ot} = \Pr(\gamma_k / \bar{\gamma}_k < \gamma_{th}) \text{ for all } k = 1 \ldots K \tag{3.6}$$

$$P_{ot} = (1 - \exp(-\gamma_{th}))^K \tag{3.7}$$

when the channel gains of all users are assumed to be i.i.d. Rayleigh fading. All users are assumed to have the same average SNR $\bar{\gamma}_k$.

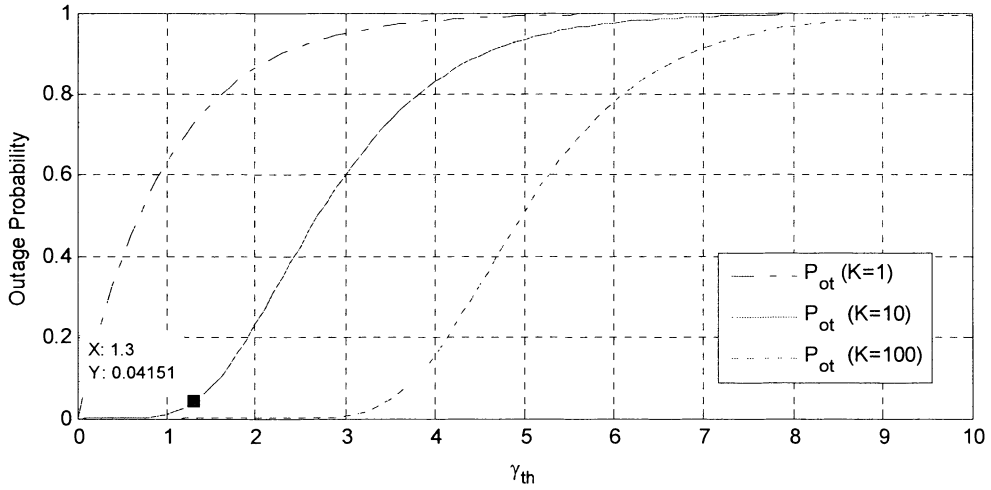


Figure 3.2 Probability of feedback outage

Figure 3.2 shows the feedback outage probability in a multiuser system with 1, 10, and 100 users employing the B-bits selective feedback.

In Figure, 3.2, the feedback outage probability increases with an increase in the value of feedback threshold. For a given value of feedback threshold, a higher value of K maintains a lower value of feedback outage. This is because the probability of none of the users having received SNR above the threshold when K = 100 is much lower than when K = 10 or when K =1.

It can be observed from Figure 3.1 that the throughput starts to drop from γ_{th} = 1.3 which is corresponding to feedback outage probability = 0.03 (3%) in Figure 3.2. For a given value of feedback outage probability, the feedback threshold values can be determined by [20]:

$$\gamma_{th} = -\log(1 - P_{ot}^{1/K}) \tag{3.8}$$

where log(·) is a natural logarithm and P_{ot} is the probability of feedback outage.

3.3.2 Effect of Feedback Threshold for One-bit Feedback

As mentioned in the previous section, the one-bit feedback scheme further reduces amount of feedback load by allowing each user to send one-bit feedback indicating

whether their received SNR is above the feedback threshold and the base station randomly allocates the resource to a user amount all users reporting the one-bit CSI feedback. In this scenario, the achievable system capacity would be more sensitive to the choice of feedback threshold as it involves the scheduling performance to allocate the resource to the best user.

Under the same simulation parameters as Figure 3.1, the achievable throughput performance of a multiuser system employing one-bit feedback scheme as a function of feedback threshold is illustrated in Figure 3.3.

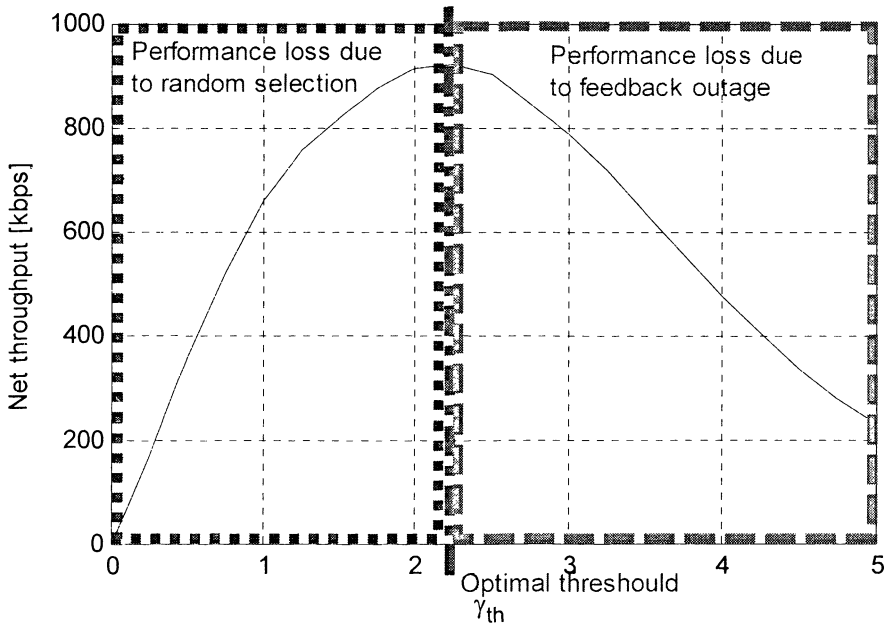


Figure 3.3 Performance of multiuser system with one-bit selective feedback

Figure 3.3 shows the achievable net throughput of a multiuser system with 10 users where one-bit feedback [22] is used. It indicates that an optimal feedback threshold exists for the one-bit feedback scheme at approximately $\gamma_{th} = 2.25$. On the left-hand side of the optimal threshold value of the feedback threshold, the performance drop is due to the inappropriate random user selection. On the right-hand side of the optimum threshold the performance loss is due to feedback outage as discussed previously.

The simulation result in Figure 3.3 demonstrates that the one-bit selective feedback technique is more sensitive to the selection of feedback threshold than B-bit selective feedback technique. Furthermore, a severe loss in the achievable net throughput can occur due to the random user selection.

More specifically, if the threshold is set too low, many users may be reporting the one-bit CSI message, and it would then be a small possibility that the base station would schedule the user having the highest received SNR in each scheduling slot. This could result in the performance loss due to inappropriate random resource allocation to a user having a weak channel condition. In contrast, if the feedback threshold is set at a larger value, the chance that the base station may randomly allocate the resource to the best user becomes higher. However, having a large value for feedback threshold is not always beneficial to the system performance as it will increase the feedback outage and the achievable capacity will accordingly drop as mentioned earlier.

3.4 Selective Feedback Using Relative SNR

The Gesbert's selective feedback and one-bit feedback techniques are based on the received SNR (i.e. received SNR). The use of received SNR leads to a problem that users located closer to the base station would be more likely to be scheduled than those located farther apart. As a result, fairness to access to the radio resources among all users cannot be maintained. This section describes some modifications that need to be made to the B-bit selective and one-bit feedback techniques so that the two techniques can be implemented with proportional fair scheduling.

3.4.1 Gesbert's Selective Feedback Using Relative SNR

To ensure that all users would have approximately the same possibility to be scheduled, relative SNR shall be used in lieu of received SNR to determine whether the users are allowed to send feedback as expressed in Equation 2.22.

Let us consider a two-user system with selective feedback threshold as shown in Figure 3.4.

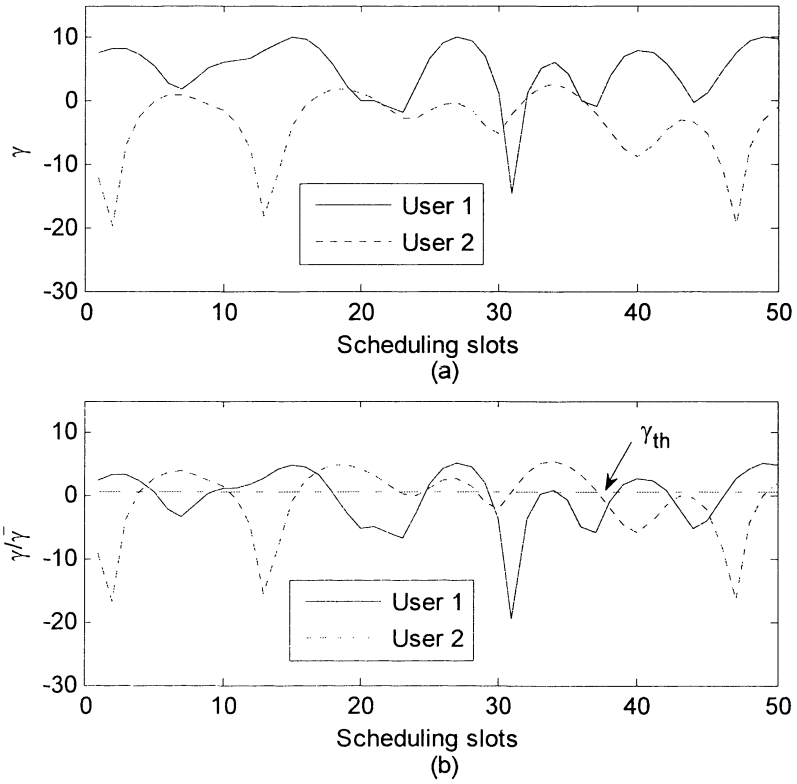


Figure 3.4 (a) Received SNR and
(b) Relative SNR

Figure 3.4(a) shows the received SNR of a system having two users where the long term average SNR ($\bar{\gamma}$) = 5dB for User 1 and -3 dB for User 2. Figure 3.1(b) shows the relative SNR ($\gamma/\bar{\gamma}$) and the feedback threshold (set as 0.7 dB). Each of the users may send feedback only if the relative SNR becomes higher than the feedback threshold. A snapshot of CSI feedback from the two users is shown in Figure 3.5.It can be seen from Figure 3.5 that User 1 and User 2 send CSI only when their relative SNR level is above the feedback threshold.

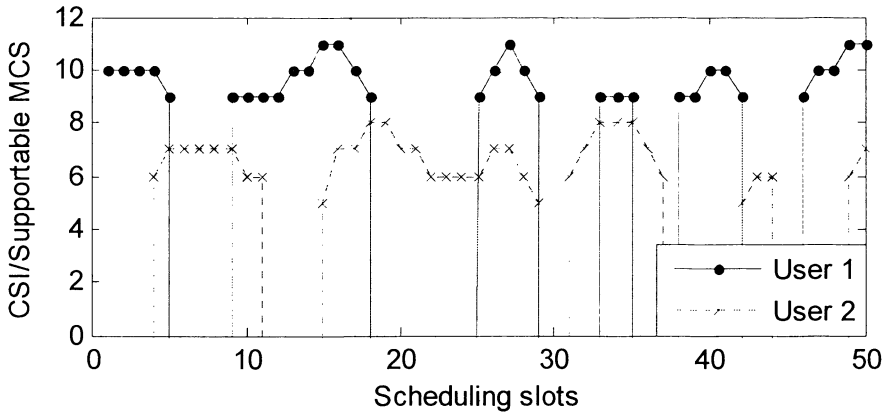


Figure 3.5 Selective CSI commands

3.4.2 One-Bit Feedback Using Relative SNR

The one-bit feedback technique uses received SNR to determine whether users should send feedback similar to the Gesbert's feedback technique.

,
Figure 3.6 shows a snapshot of MCS states in a system with two users employing one-bit feedback technique based on the relative SNR in Figure 3.4(b)

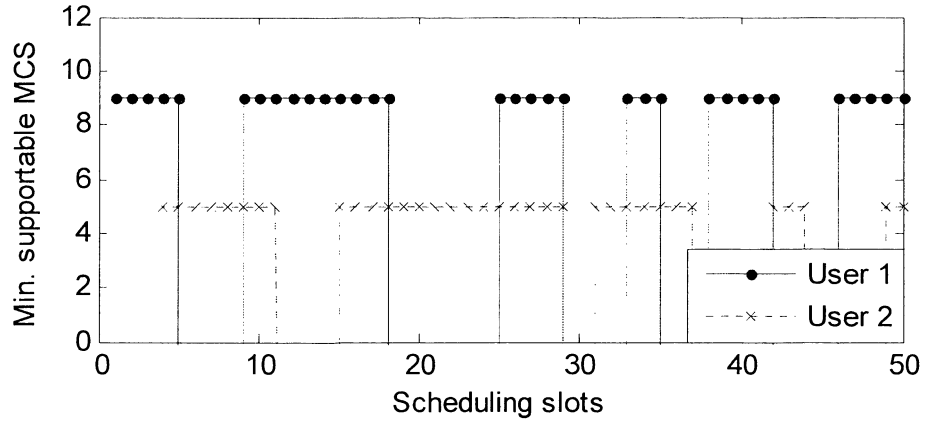


Figure 3.6 One-bit CSI commands

Each user in Figure 3.6 sends a one-bit CSI message to indicate if the relative SNR is higher than the threshold (see Figure 3.4(b)). Note that User 1 and User 2 in Figure 3.6 send one-bit CSI to the base station. The base station then determines a supportable MCS based on the following criteria:

$$MCS_{1-bit}(k) = MCS(\gamma_k') \tag{3.9}$$

where γ'_k is the received SNR that the base station interprets from the average SNR and the feedback threshold as expressed below:

$$\gamma_k' > \bar{\gamma}_k \cdot \gamma_{th} \tag{3.10a}$$

or

$$\gamma_k' = \bar{\gamma}_k \cdot \gamma_{th} \tag{3.10b}$$

when the lower limit of the interpreted received SNR is used.

Equation 3.10b is based on the fact that a user would send feedback if and only if the relative SNR is above the feedback threshold. The value of the feedback threshold is known to the base station while the average SNR for each user needs to be reported to the base station. The average SNR changes at a significantly slower rate compared with the received SNR (which is an instantaneous value of SNR). There are three options for the base station to determine the average SNR of each user so that the received SNR can be calculated:

i) The average SNR can be updated periodically in a much longer update interval than the scheduling slot
ii) The average SNR can be updated when the supportable MCS corresponding to the average SNR changes (event triggered report).
iii) The system can determine the average SNR from HARQ sequence

It is known from the previous example associated with Figure 3.4 that the average SNR of User 1 is 5 dB, and the feedback threshold is 0.7 dB. If the base station receives a feedback command from User 1, it will interpret that the user has the received SNR of greater than 5 + 0.7 = 5.7 dB. This SNR value can then be mapped to an appropriate MCS. In this case, the base station will interpret that this user can support 1288.8 kbps (MCS 9 in Table 2.1).

3.5 Proposed Feedback Algorithm

The proposed feedback technique is based on a similar feedback threshold concept as discussed in the previous sections. It is developed based on an observation that

the MCS values obtained by the B-bits selective feedback scheme are always higher than or equal to the MCS value that the base station can calculate from the one-bit feedback. The new technique aims to bridge the performance gap between the one-bit feedback and selective feedback schemes in terms of achievable net throughput and amount of feedback load.

3.5.1 Graphical Representation of Proposed Feedback Algorithm

To illustrate the behaviour of the proposed feedback reduction technique, let us consider the supportable MCS modes based on one-bit feedback and selective feedback in Figure 3.7.

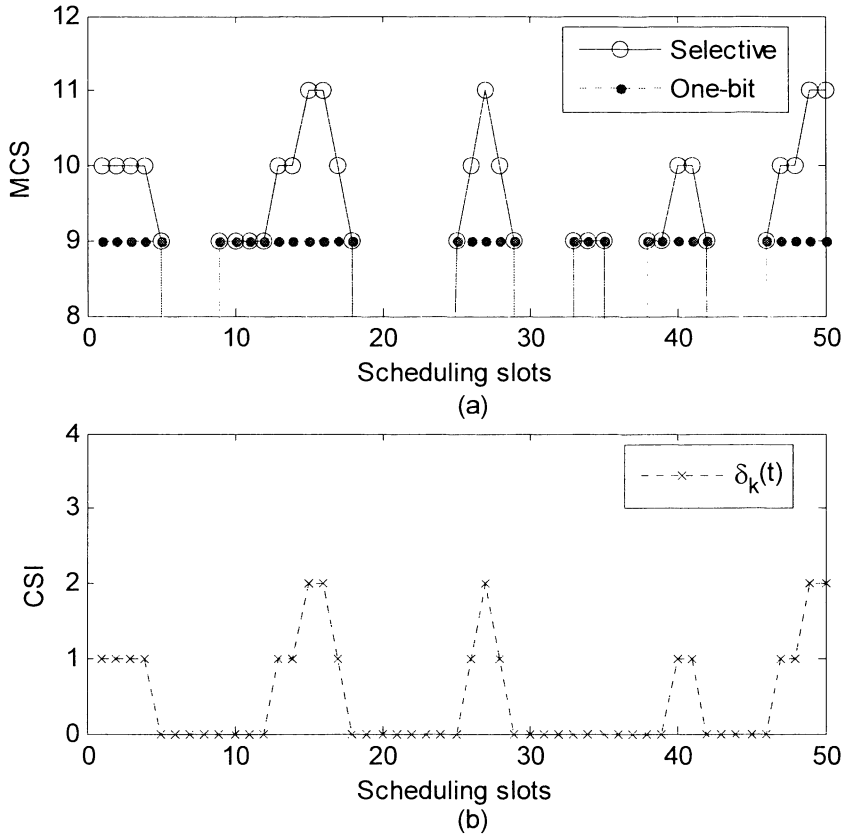


Figure 3.7 (a) Supportable MCS, and (b) Difference in MCS, $\delta_k(t)$

Figure 3.7 (a) shows the supportable MCS modes when User 1 is sending one-bit feedback, and B-bits selective feedback, respectively. Figure 3.7 (b) shows the difference, $\delta_k(t)$, between the supportable MCS modes when a user sending B bits

selective feedback and one-bit feedback. It can be observed that between the scheduling slots 12 and 18 where the received SNR is reaching a peak (see Figure 3.4), the supportable MCSs are relatively high (see Figure 3.7(a)). However, the one-bit feedback cannot capture this peak, and thus it does not report to the base statioin to increase MCS mode during this peak. As a result, the achievable capacity of one-bit feedback technique is less than the achievable capacity with full- CSI.

To eliminate the capacity loss caused by the use of one-bit feedback while minimising the feedback load, a new feedback scheme is proposed in this section. With this proposed feedback reduction scheme, a mobile user whose relative SNR is higher than an optimally selected threshold sends a two-bit CSI command to report the difference between the supportable MCS modes of the B-bits selective technique and one-bit feedback technique .

3.5.2 Mathematical Representation of Proposed Feedback Algorithm

The MCS difference, $\delta_k(t)$, can be expressed as follows:

$$\delta_k = \min(MCS(\gamma_k) - MCS(\bar{\gamma}_k \cdot \gamma_{th}), \delta_{\max}) \qquad (3.11)^6$$

where $MCS(\gamma_k)$ is defined as a MCS mapping function to map between the k^{th} user's received SNR γ_k, and an appropriate MCS mode based on Table 2.1. δ_{max} is the maximum reportable MCS difference that can be reported by the δ_k command. If . $\delta_k(t)$ consists of 2 bits, it can capture $2^2 = 4$ levels in the MCS difference.

An example of CSI configuration with 2 bits for the proposed feedback reduction technique can be described as:

$$CSI_k = \begin{cases} 00 & \delta_k = 0 \\ 01 & \delta_k = 1 \\ 10 & \delta_k = 2 \\ 11 & \delta_k = 3 \end{cases} \qquad (3.12).$$

Then, the base station determines the current supportable MCS of the k^{th} user using:

[6] Hereafter, notation for time (t) is omitted for simplification.

$$MCS_k = MCS(\bar{\gamma}_k \cdot \gamma_{th}) + \delta_k \quad (3.13)$$

The fundamental ideas of the proposed scheme consist of the following items:

a) Consider the terms on the right hand side of Equation 3.13, the system knows that the received SNR of users sending feedback must be greater than or equal to the threshold value, and thus $MCS(\bar{\gamma}_k \cdot \gamma_{th})$ is the minimum supportable MCS when a user is sending feedback.

b) CSI includes extra information to report the difference between $MCS(\bar{\gamma}_k \cdot \gamma_{th})$ and the supportable MCS. This piece of information is intended to capture the difference in the actual supportable MCS and the minimum supportable MCS from Item a) above, hereafter MCS difference. However, the proposed feedback reduction technique can report the difference in MCS up to the value of δ_{max} only.

In the next section, an optimal value of δ_{max} is investigated.

3.5.3 Determining an Optimal Value for δ_{max}

It is highly probable that the difference between the supportable MCS mapped from the received SNR and that mapped from the interpreted SNR in Equation 3.10b to be less than δ_{max}. To further investigate this, the probability of such event, P_{δ}, is set:

$$P_{\delta}(\delta_{\max}) = \Pr\left(\left(MCS(\gamma_k) - MCS(\bar{\gamma}_k \cdot \gamma_{th})\right) > \delta_{\max}\right) \quad (3.14)$$

The probability is Equation 3.14 can be rearranged as follows:

By definition, $\delta_{\max} = (MCS(\bar{\gamma}_k \cdot \gamma_{th}) + \delta_{\max}) - MCS(\bar{\gamma}_k \cdot \gamma_{th})$, Equation 3.14 can be rewritten as:

$$P_{\delta}(\delta_{\max}) = Pr\left(\left(MCS(\gamma_k) - MCS(\bar{\gamma}_k \cdot \gamma_{th})\right) > (MCS(\bar{\gamma}_k \cdot \gamma_{th}) + \delta_{\max}) - MCS(\bar{\gamma}_k \cdot \gamma_{th})\right)$$

As $MCS(\gamma_k)$ is an increasing function of γ_k, Equation 3.15 can be simplified as:

$$P_\delta(\delta_{\max}) = Pr\left(\gamma_k - \bar{\gamma}_k \cdot \gamma_{th} > \Gamma_{MCS_{k+\delta_{\max}}} - \Gamma_{MCS_k}\right) \quad (3.16a)$$

or

$$P_\delta(\delta_{\max}) = Pr\left(\gamma_k > \bar{\gamma}_k \cdot \gamma_{th} + \left(\Gamma_{MCS_{k+\delta_{\max}}} - \Gamma_{MCS_k}\right)\right) \quad (3.16b)$$

where

MCS_k is the minimum supportable MCS of user k i.e. $MCS(\bar{\gamma}_k \cdot \gamma_{th})$,

Γ_{MCS} is the SNR quantisation value of $MCS(\bar{\gamma}_k \cdot \gamma_{th})$ (as per Table 2.2)'

and $\Gamma_{MCS+\delta max}$ is SNR quantisation value of $MCS(\bar{\gamma}_k \cdot \gamma_{th}) + \delta_{max}$.

By dividing all terms in Equation 3.16b by $\bar{\gamma}_k$, the following equation can be obtained:

$$P_\delta(\delta_{\max}) = Pr\left(\frac{\gamma_k}{\bar{\gamma}_k} > \gamma_{th} + \frac{\left(\Gamma_{MCS_{k+\delta_{\max}}} - \Gamma_{MCS_k}\right)}{\bar{\gamma}_k}\right) \quad (3.17)$$

For Rayleigh fading channel, the probability that the SNR γ is greater than an arbitrary value x is:

$$Pr(\gamma > x) = exp(-x) \quad (3.18)$$

Thus, From Equation 3.17 and 3.18, Equation 3.14 can be reformulated as:

$$P_\delta(\delta_{\max}) == \exp\left(-\gamma_{th} - \frac{\left(\Gamma_{MCS_{k+\delta_{\max}}} - \Gamma_{MCS_k}\right)}{\bar{\gamma}_k}\right) \quad (3.19)$$

Equation 3.19 is a discrete function with non-uniform SNR-to-MCS mapping quantisation levels. Mathematically solving this equation is beyond the scope of this thesis. Thus, Equation 3.19 is evaluated using simulations and the result is given for γ_{th}=1, and 2.25 in Figure 3.8. The average SNR is chosen to be -2.2 dB based on some initial simulations as it is the worst case that MCS difference may occur.

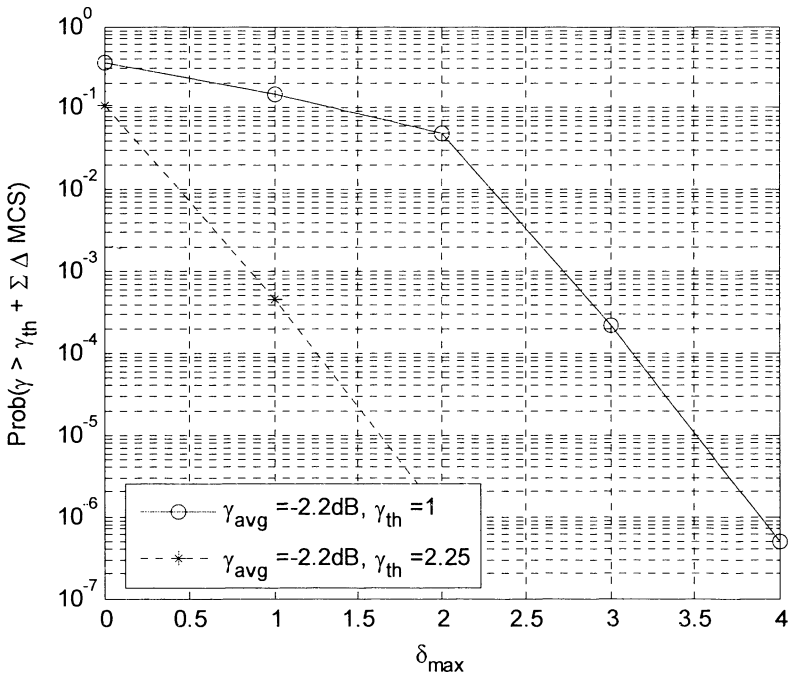


Figure 3.8 Solution of Equation 3.19

The result in Figure 3.8 shows that the probability that the MSC mapped from the received SNR exceeds the MCS mapped from the interpreted SNR in Equation 3.10b dramatically reduces when δ_{max} increases..

Furthermore, it can be observed from Figure 3.8 that, the probability that the MCS difference may exceed $\delta_{max} = 3$ is less than 10^{-3}. This probability indicates that the proposed feedback reduction technique with a small value of $\delta_{max} = 3$ is sufficient to report supportable MCS with more than 99.9% accuracy.

Hence, from above explanations, the maximum of feedback bits per CSI command is limited to two bits which can represents a maximum of 2^2 -1 = 3 different additional MCS levels i.e. $\delta_{max} = 3$ and $\delta_k(t) \in [0, \delta_{max}]$.

Once the base station receives a feedback command, $CSI_k(t)$ ($CSI_k(t) \equiv \delta_k(t)$ for the proposed two-bit feedback scheme), from the k^{th} user, it knows the minimum supportable MCS is $MCS(\bar{\gamma} \cdot \gamma_{th})$ as discussed in the previous subsections.

Given that the system considered in this chapter has 12 MCS modes (4 bits required), the proposed scheme requires approximately half of the feedback load compared with the selective scheme proposed in [20]. Although our scheme requires nearly double the feedback load compared with the one-bit feedback presented in [21][22], this scheme achieves significantly higher net throughput than that achieved by the one-bit scheme. More details is presented in Section 3.6 below.

3.6 Simulation Results

This section presents simulation results in various aspects of the proposed feedback reduction technique namely: Achievable net throughput, Optimum thresholds, Feedback load, and MCS update error probability. The effect of user movements to each link between user and the BS is modelled as a correlated Rayleigh fading with Doppler frequency of 0.067 Hz (10m/s at 2GHz). Users' SNR values are generated from uniform distributed. No explicit mobility models used. The simulation parameters are summarised in table below:

Table 3.1 Summary of Simulation Parameters (Monte Carlo simulation)

Description	Parameters Used
Number of users, K	From 0 to 100
Channel model	Rayleigh fading with Jake's power spectrum density (10m/s at 2 GHz)
Scheduling algorithm	Proportional fair
Adaptive MCS	As per Table 2.2
Average SNR	Uniformly distributed from -5 to 5 dB
Scheduling slot	1 ms
Traffic model	Full buffered
Feedback channel	Error free and delay free.

The optimum threshold, $\hat{\gamma}_{th}$, is the threshold that maximises the net threshold for a given number of users k.

$$\hat{\gamma}_{th}(k) = \arg\max_{\gamma_{th}}(C'(\gamma_{th})) \tag{3.20}$$

where C' represents the net throughput in kbps.

For the selective feedback, a threshold that maximises the net throughput will be referred to as an optimal threshold. For selective feedback, a threshold that results in a 2% capacity loss compared with the full capacity is chosen. Note that the optimal feedback threshold is a function of number of users. It is not necessary to update optimal feedback thresholds every scheduling slot as long as the number of users does not change.

The optimal thresholds for the three feedback reduction schemes for different number of users are given in Table 3.2 (see also Equation 3.20). It can be seen from Table 3.2 that the values of optimum thresholds increase with an increase of the number of users.

Table 3.2: Optimal Feedback Thresholds

	Number of users, K					
	2	**5**	**10**	**20**	**50**	**100**
Selective feedback	0.70	0.90	1.25	2.25	3.25	4.00
Proposed feedback	0.30	0.70	1.00	2.25	3.25	4.00
One-bit feedback	1.10	1.75	2.25	2.50	3.50	4.25

The feedback load for each scheme is defined as the amount of feedback bits from users sending feedback in each scheduling slot normalised by the total number of users [20]. The normalised feedback load for one-bit can be computed from

$$\overline{F}_{1-bit} = \frac{E[K_{\gamma_{th1}}(t)]}{K} \tag{3.21}$$

where $E[x]$ is the expectation value or mean value of a variable x, and $K_{\gamma th}(t)$ is the number of feedback users (out of K users) at time t when the threshold γ_{th1} is set[7]. Similarly, the normalised feedback load for selective and our proposed scheme can be expressed as:

$$\overline{F}_{selective} = \frac{4 \cdot E[K_{\gamma_{th}}(t)]}{K} \tag{3.22}$$

and

$$\overline{F}_{proposed} = \frac{2 \cdot E[K_{\gamma_{th}}(t)]}{K}. \tag{3.23}$$

In the Rayleigh fading scenario, Equation 3.21 becomes

$$\overline{F}_{1-bit} = \Pr(\gamma > \gamma_{th}) = 1 - (1 - e^{-\gamma_{th}/\bar{\gamma}}) = e^{-\gamma_{th}/\bar{\gamma}} \tag{3.24}$$

Therefore, the average of the normalised feedback load for the one-bit feedback reduction technique, FB_{1bit}, can be computed by:

$$FB_{1bit} = \int_{-\infty}^{\infty} \overline{F}_{1-bit} p_{\gamma} d\bar{\gamma} = \int_{\gamma_{\min}}^{\gamma_{\max}} \overline{F}_{1-bit} p_{\gamma} d\bar{\gamma} \tag{3.25}$$

where p_{γ} is the probability distribution function of the long-term average SNR, which follows uniform distribution. The normalised feedback for the selective feedback, $FB_{selective}$, and the proposed feedback, $FB_{proposed}$, can then be computed by:

$$FB_{selective} = 4 \cdot FB_{1bit} \tag{3.26}$$

and

$$FB_{proposed} = 2 \cdot FB_{1bit} \tag{3.27},$$

respectively.

The proposed feedback technique is based on the fact that the CSI commands sent by the selective feedback techniques have higher MCS values than or equal to MCS interpreted by the one-bit technique. Therefore, errors may occur if the difference between the supportable MCS values of the full-CSI and the one-bit technique are

[7] The value of optimal feedback threshold (γ_{th}) and the number of users sending feedback ($K_{\gamma th}$) for each feedback reduction scheme are different. The values in Equation 3.21 to 3.24 shall be of the corresponding schemes. For simplification, same notifications are used for all schemes throughout these equations.

greater than the maximum value δ_{max}. With the proposed feedback reduction technique, one CSI message can represent at most four levels in CSI difference since the amount of feedback per CSI is limited to two bits (4 stages: 0, 1, 2 and 3). Therefore, MCS update error will occur if the difference between the supportable MCS modes of the one-bit and full-CSI schemes is greater than 3. The probability of the MCS update error, $Pe_k(t)$, can be expressed by the following error probability:

$$Pe_k(t) = \Pr(MCS_k(t) - MCS(\bar{\gamma} \cdot \gamma_{th}) > \delta_{\max} \; when \; \delta_{\max} = 3) \qquad (3.28)$$

3.6.1 Achievable Net Throughput

In this section, the performance of the proposed scheme in terms of achievable net throughput is studied and compared with the net throughput obtained from full-CSI, B-bits selective feedback [20], and one-bit feedback [21][22]. For the proposed scheme and one-bit feedback scheme, it is assumed that the base station knows the average SNR of each user by one of the three techniques discussed previously.

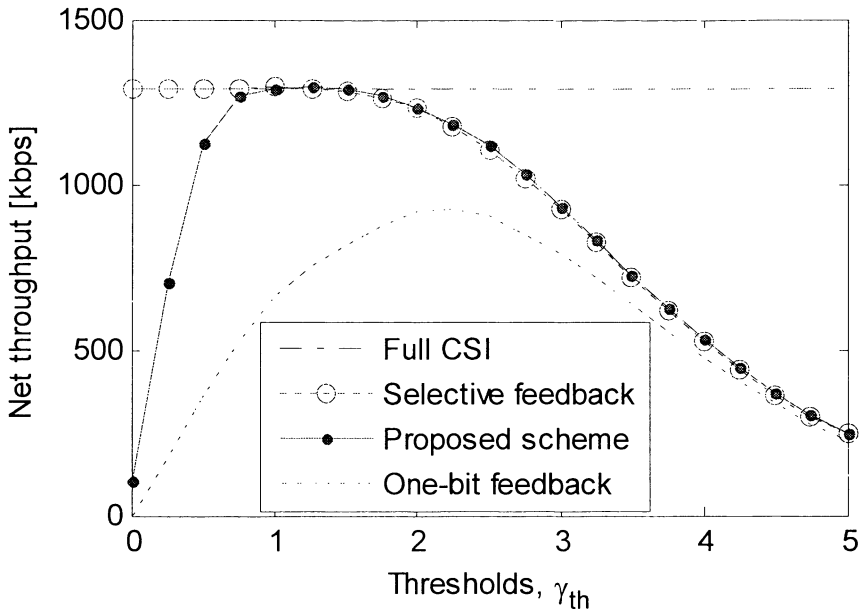


Figure 3.9 Net throughput versus feedback threshold

Figure 3.9 shows the net throughput in kbps of the four schemes versus the feedback threshold in when there are 10 users moving at a speed of 10 m/s. It can

be observed that the proposed scheme can achieve similar net throughput to full-CSI (approx. 1295 kbps), provided that an optimal threshold is selected (1.25 in Figure 3.9). An inappropriate selection of feedback threshold has a significant effect on the system throughput. In particular, if the threshold was set too small, the difference between the supportable MCS and that interpreted by the base station may become larger than δ_{max} resulting in MCS update error (See Section 3.6.4 for details). On the other hand, if the threshold is set too large, the probability that a user may have a relative SNR greater than the threshold becomes smaller, resulting in a feedback outage in which none of the users sends a CSI report and thus the base station cannot allocate radio resource to any user.

From Figure 3.9, the one-bit feedback can achieve a throughput of approx. 921 kbps when the optimal threshold is selected when the proposed scheme can achieve around 1295 kbps. This is corresponding to a throughput degradation of 1295 – 921 = 374 kbps or 28.9%.

3.6.2 Achievable Net throughput with Optimal Feedback Threshold

The achievable net throughput with optimally selected thresholds versus number of users is plotted in the Figure 3.10. As expected, the net throughput in Figure 3.10 rises as the number of users increase due to multiuser diversity as described in Section 1.2.2. The increasing rate reduces at a higher number of users because the rate improves in terms of log(log(k)) as derived in [9].

Figure 3.10 demonstrates that one-bit feedback has the lowest system throughput while the proposed scheme achieves approximately the same throughput as the selective feedback. However, the proposed scheme requires approximately half of the feedback bits required for the selective feedback scheme. This will be discussed further in the next section.

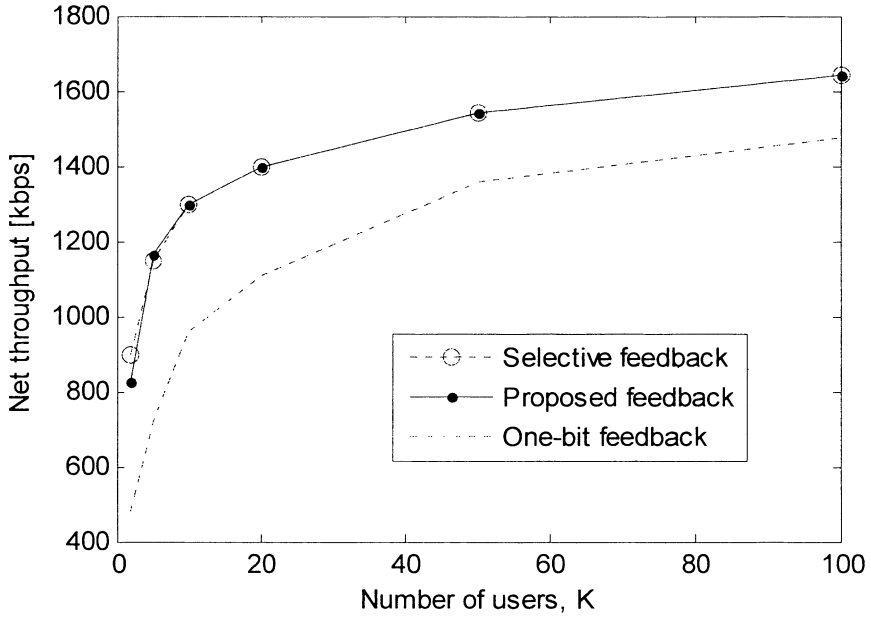


Figure 3.10 Net throughput versus number of users

The achievable net throughput of the proposed technique normalised by the throughput of full CSI is given in Table 3.3. The achievable net throughput of one-bit feedback is given for comparison.

Table 3.3 Achievable net throughput of the proposed technique

Number of user, K	**2**	**5**	**10**	**20**	**50**	**100**	**Average**
Throughput of proposed normalised by full CSI	88.9%	98.4%	98.5%	98.6%	98.7%	98.8%	97.0%
Throughput one-bit normalised by full CSI	52.2%	56.0%	70.5%	77.5%	84.1%	89.7%	71.7%
% Improvement of proposed compared with one-bit	36.7%	42.4%	28.0%	21.1%	14.6%	9.1%	25.3%

From Table 3.3, it can be observed that the proposed technique obtains approximately 97% of the full-CSI net throughput. The proposed technique

provides around 25.3% improvement in the net throughput when compared with one-bit feedback technique

3.6.3 Feedback Load

This section presents the reduction in feedback load of the proposed scheme in comparison with other feedback reduction schemes.

For the purpose of feedback load comparison, it is assumed that the supportable MCS mapped from the average SNR is updated periodically every 20 scheduling slots which would contribute to an increase of approximately 5% in the total feedback load required in the proposed scheme. The number 20 was selected for the purpose of feedback load comparison. An update of once per 20 slots is approximately an additional 1/20 = 0.05 or 5% of feedback load increase. Any arbitrary number greater than 3 (33% increase) would show that the proposed feedback reduction techniques requires less feedback load than the Selective feedback as the Selective feedback technique requires 4 bits per CSI message while the proposed technique requires 2 bits (50% less).

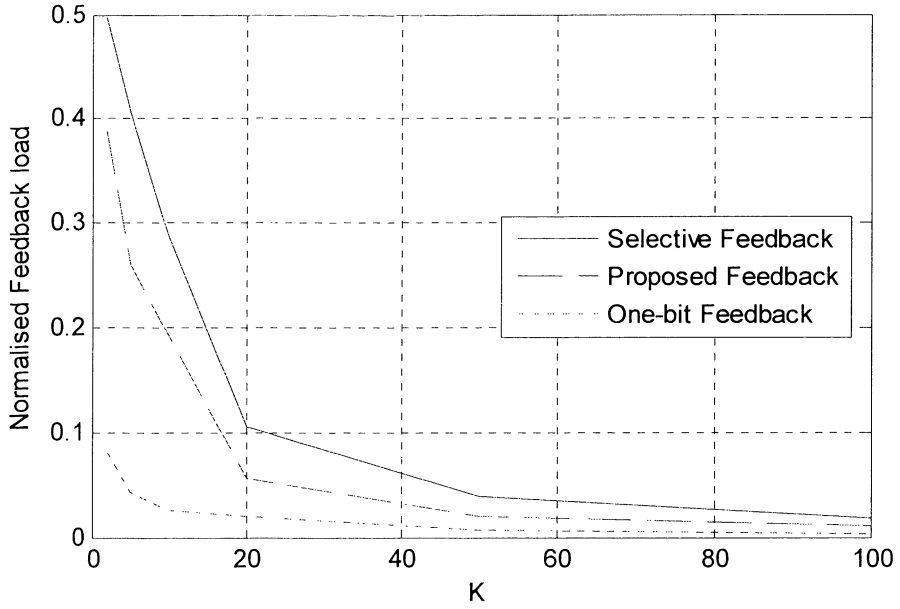


Figure 3.11 Feedback load versus number of users with optimal thresholds

Figure 3.11 shows the normalised feedback load of various feedback reduction schemes. The feedback loads are normalised with the full-feedback load. The feedback load is calculated based on the optimal feedback threshold values of each corresponding feedback technique which are different for each technique to each other.

The results in Figure 3.11 demonstrate that the proposed scheme requires less amount of feedback load compared with the selective scheme. It can be observed that normalised feedback load reduces significantly when the number of users grows. This is because the optimal threshold is set a larger value when the number of users becomes greater and only users whose relative SNR is sufficiently high are allowed to send CSI feedback. To further illustrate this, the average numbers of users sending feedback for the three feedback schemes are plotted in Figure 3.12 below:

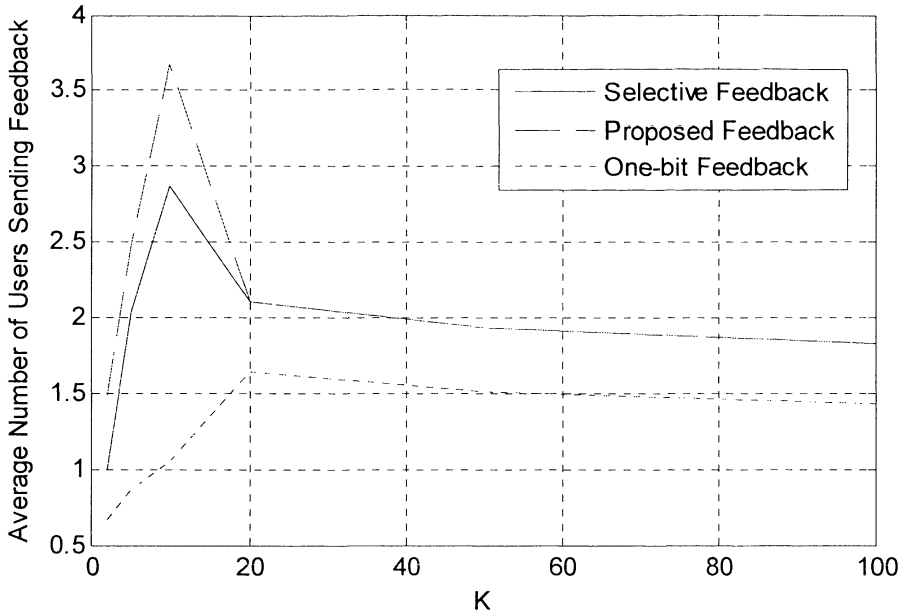


Figure 3.12 Feedback load comparison

In Figure 3.12 when the number of users is relatively low (i.e. less than 20 users for the proposed and selective feedback techniques) the value of optimal feedback threshold is relatively small and thus more users are sending feedback on average. When the number of users is high, the average number of users sending feedback

are nearly constant. This result suggests that the optimal feedback thresholds that have been set to maximise the achievable throughput can maintain the number of feedback users remaining constant.

More details in evaluating optimal/suboptimal feedback thresholds will be studied in Chapter 4.

3.6.4 MCS Update Error Probability

The error probability is plotted versus the feedback thresholds in Figure 3.13.

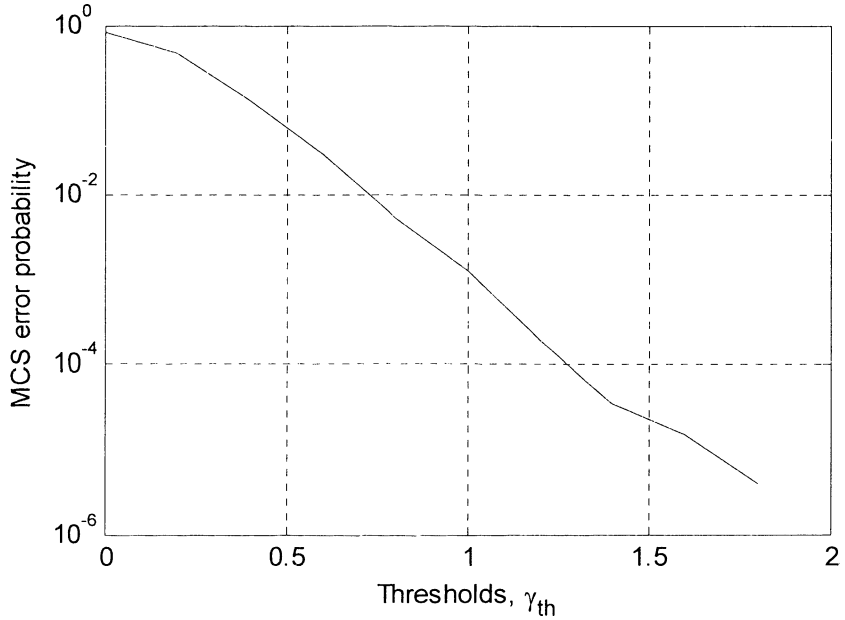


Figure 3.13 MCS error probability versus thresholds

Figure 3.13 shows the MCS update error probability at the base station. Referring to Table 3.2, the optimum threshold for 5 users is 0.7 which corresponds to an MCS error probability of less than 0.01 (or 99% accuracy). For 10 users the optimum threshold is 1 which corresponds to 0.001 error probability (or 99.99% accuracy). It can be observed that the error decreases rapidly as the value of threshold increases. It can be expected that the MCS error probability would increase when the value of feedback threshold grows because only users whose supportable MCS is high would send feedback and thus the proposed feedback reduction technique can capture users reaching their peak received SNR. Therefore, the proposed scheme

achieves nearly the same performance as the selective feedback with approximately 50% reduction in feedback load.

3.7 Summary

In this chapter, feedback reduction techniques proposed in the literature are reviewed and briefly discussed in Section 3.2. The effect of feedback threshold of the performance of B-bits selective feedback and one-bit feedback, and the modified version for the two feedback schemes are studied in Section 3.3 and 3.4 respectively. It is demonstrated through simulation snapshots that the one-bit feedback proposed cannot make full use of the high-peaks of the channel conditions resulting in throughput loss.

In Section 3.5, a new feedback reduction technique that sends two or more bits for a CSI command per scheduling slot is presented and evaluated. It is shown in Section 3.5. that the probability that the difference between instantaneous MCS and MCS from average SNR exceeding a predefined value (i.e. δ_{max}) approaches zero for a small value of δ_{max} (i.e. δ_{max} =3) . Simulation results in Section 3.6 verify that the net throughput obtained by the proposed scheme with two-bit feedback outperforms that obtained by the one-bit feedback scheme by 25.3% on average. Approximately the same net throughput as selective feedback can be achieved with around 40% feedback reduction. The accuracy of SNR-to-MCS mapping which is obtained from CSI of the proposed feedback technique with δ_{max} =3 exceeds 99%.

Chapter 4

Link Adaptation with Multiple Feedback Thresholds

This chapter deals with an evaluation of a feedback reduction technique with multiple feedback thresholds. Section 4.1 presents an overview of feedback reduction techniques in the literature that implemented multiple feedback thresholds. The achievable sum rate of multiuser system using single feedback threshold is evaluated in Section 4.2. Section 4.3 formulates the achievable sum rate of multiuser system with multiple feedback thresholds. The numerical results obtained from the proposed formulation are presented in Section 4.4. Then, in Section 4.5, suboptimal thresholds for a multiuser system with single-feedback and two-feedback thresholds are evaluated. Section 4.6 concludes this chapter.

4.1 Overview of Feedback Techniques with Multiple Thresholds

Feedback reduction techniques in the literature that implemented multiple feedback thresholds are reviewed in this section.

4.1.1 Hassenl's Multiple Feedback Thresholds Technique

In [26], Hassenl et al presented a multiple threshold feedback reduction algorithm where the users are first probed with the highest threshold. If received SNR of the users are below this threshold then the threshold value is successively lowered until one or more users send a CSI message. Although this technique claims to guarantee zero feedback outage, it needs to perform polling until a user responds within a given feedback time limit. Another requirement of the feedback threshold technique in [26] is that it involves relatively large number of feedback thresholds (typically 10 in the simulation) which would make finding the optimal threshold for all users and practical implementation more challenging.

4.1.2 Nam's Multiple Feedback Thresholds Technique

In [27], Nam et al presented a new multiuser scheduling with a specific value of feedback threshold for each user. The system implementing Nam's technique allocates a different threshold to each user based on long term average SNR. Users having SNR higher than the optimally adjusted thresholds are allowed to send a feedback message. Nam's feedback algorithm improves the achievable capacity when compared with using fixed threshold. However, it involves optimising individual threshold for each user which would be considered as an extra computational overhead required to implement this algorithm in practice.

4.1.3 So's Multiple Feedback Thresholds Technique

In [28], a new opportunistic feedback was presented for multiuser systems with proportional fair and maximum throughout scheduling. The proposed feedback protocol separates a contention-based feedback period into two sub-periods each of which is assigned a different feedback threshold value. Users particulate in these two minislots with different probability. Despite the performance improvement gained from the proposed protocol compared with a system with one feedback period, it did not utilise the multiple feedback thresholds to reduce the amount of feedback that each user would send when the received SNR exceeds the threshold.

It can be observed that the previous works in the literature studied different aspects of the use of multiple feedback thresholds for multiuser systems with opportunistic scheduling.

4.2 Achievable Throughput with Single Feedback Threshold

Somekh et al [22] analysed a multiuser system with opportunistic scheduling. This work forms the basis for work proposed in Section 4.3.

The base station (or the scheduler) broadcasts a feedback threshold α to all the users. Each user whose received SNR is above the threshold sends a CSI containing

bit '1' to the base station. The base station randomly selects a user to be served based on the received one-bit information. In this system, the radio channel is active only if at least one user among K users has the received SNR above the threshold α. The closed-form expression of the sum-rate, C, has been discussed in Section 3.2 and provided again here for quick reference [22]

$$C(\alpha, K, P) = Q(\alpha, K) \cdot \chi(\alpha, P) \tag{4.1}$$

where

$$Q(\alpha, K) = (1 - (1 - e^{-\alpha})^K) \tag{4.2a}$$

$$\chi(\alpha, P) = \log(1 + \alpha P) + e^{\left(\frac{1+\alpha P}{P}\right)} \cdot Ei\left(\frac{1+\alpha P}{P}\right) \tag{4.2b}$$

$Ei(x) = \int_x^\infty e^{-t}/t\, dt$ is the exponential integral function and *log(x)* is a natural logarithm of a variable x. α represents the feedback threshold and P is the average SNR.

In this chapter, the symbol α is used to represent the value of feedback threshold for the sake of notation simplification and being consistent with the original work [22] which is the fundamental of the analysis in this chapter. In other chapters in this thesis, the value of feedback threshold is represented by γ_{th}.

Similarly, symbol P is used in lieu of $\bar{\gamma}$ in this chapter only.

$Q(\alpha,K)$ is the probability that at least one user has received SNR above the threshold α, $\Pr(\gamma_m > \alpha)$ where $\gamma_m = \max_k \gamma_k$.

$$Q(\alpha,K) = 1 - P_{ot} \tag{4.3}$$

where the feedback outage probability, P_{ot}, is as described in Chapter 3.

$\chi(\alpha,P)$ is referred to as the achievable spectral efficiency of a scheduled user over a scheduling slot, in which fading is assumed to be constant

$$\chi(\alpha, P) = E_{\gamma_\alpha}\{\log(1 + P\gamma_\alpha)\} \tag{4.4}$$

where $E\{x\}$ is the mean value of random variable x, and $\gamma_\alpha \in \{\gamma_k \mid \gamma_k > \alpha\}$; $k=1,..,K$.

The achievable sum-rate of the one-bit feedback as a function of feedback threshold, α, for $K = 2$, 10, 20, 50 and 100 is illustrated in Figure 4.1.

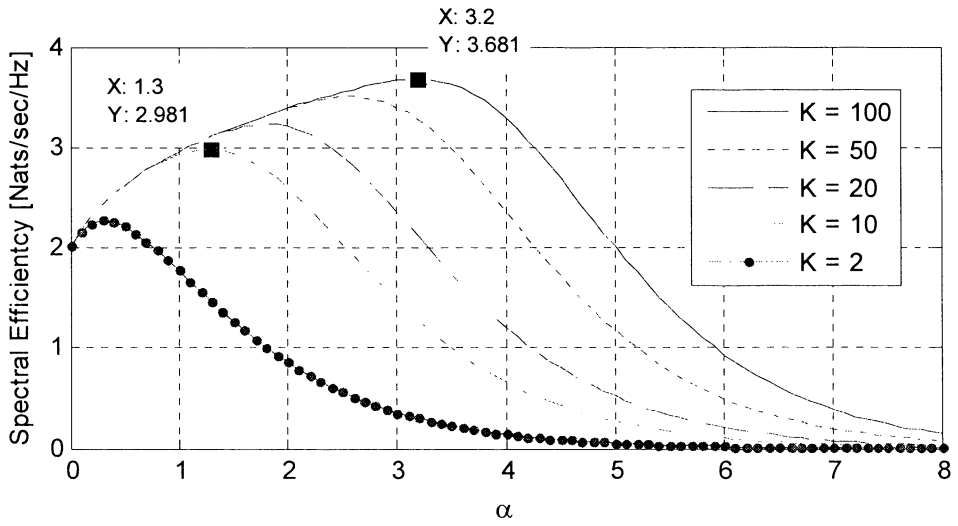


Figure 4.1 Achievable sum-rate of system with one-bit feedback with various K

The spectral efficiency (also referred to as achievable sum-rate) of the one-bit feedback, $C(\alpha,K,P)$, in Equation 4.1 as a function of α for K = 2, 10, 20, 50 and 100 is plotted in Figure 4.1. P is a constant at 10dB..

Figure 4.1 shows that the performance of a multiuser system with one-bit feedback scheme is sensitive to the choice of feedback threshold. Furthermore, it demonstrates that the optimal value of feedback threshold increases when the number of active users increases. For example, when K = 10, the optimal threshold is at α = 1.3 while when K = 100, the optimal threshold is at α = 3.2. As discussed in Chapter 2, the optimal feedback thresholds are the threshold values that maximum the achievable sum rate which can be numerically evaluated as per

Equation 3.6. The relationship between the optimal threshold and number of active users is shown in Figure 4.2.

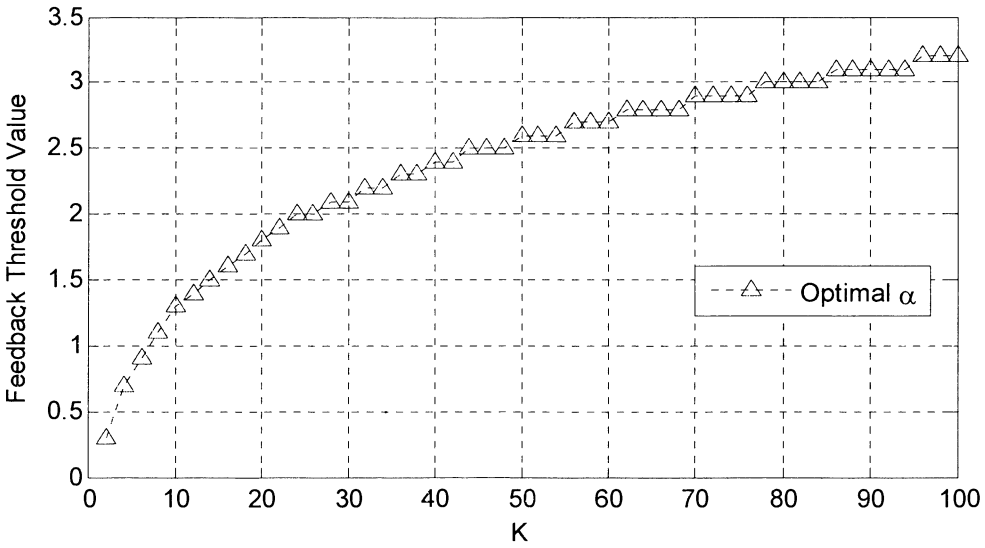


Figure 4.2 The optimal feedback thresholds vs number of active users

From Figure 4.2, it can be observed that the value of optimal threshold tends to increase when the number of active users increases.

The lower-bound capacity can be obtained by replacing the conditioned fading power γ_α in Equation 4.3 by its lowest value α [22].

$$\chi_L(\alpha,P) = log(1+P\cdot\alpha). \qquad (4.5).$$

This yields the lower-bound capacity

$$C_L(\alpha,K,P) = Q(\alpha,K)\cdot\chi_L(\alpha,P) \qquad (4.6)$$

The achievable sum-rate with optimal feedback thresholds ($C(\alpha_{op},K,P)$ when α_{op} is the optimal threshold as shown in Figure 4.2) is plotted against the achievable lower bound (Equation 4.6) in Figure 4.3:

Figure 4.3 compares the achievable sum-rate using Equation 4.1 and its lower performance bound using Equation 4.6. For the purpose of comparison, the

achievable sum-rate for the full-CSI case when the exact CSI of all users is known to the base station is plotted on the graph. The performance gap between the three techniques is noticeable. The proposed scheme with multiple feedback thresholds aims at minimising this gap while minimising the feedback load.

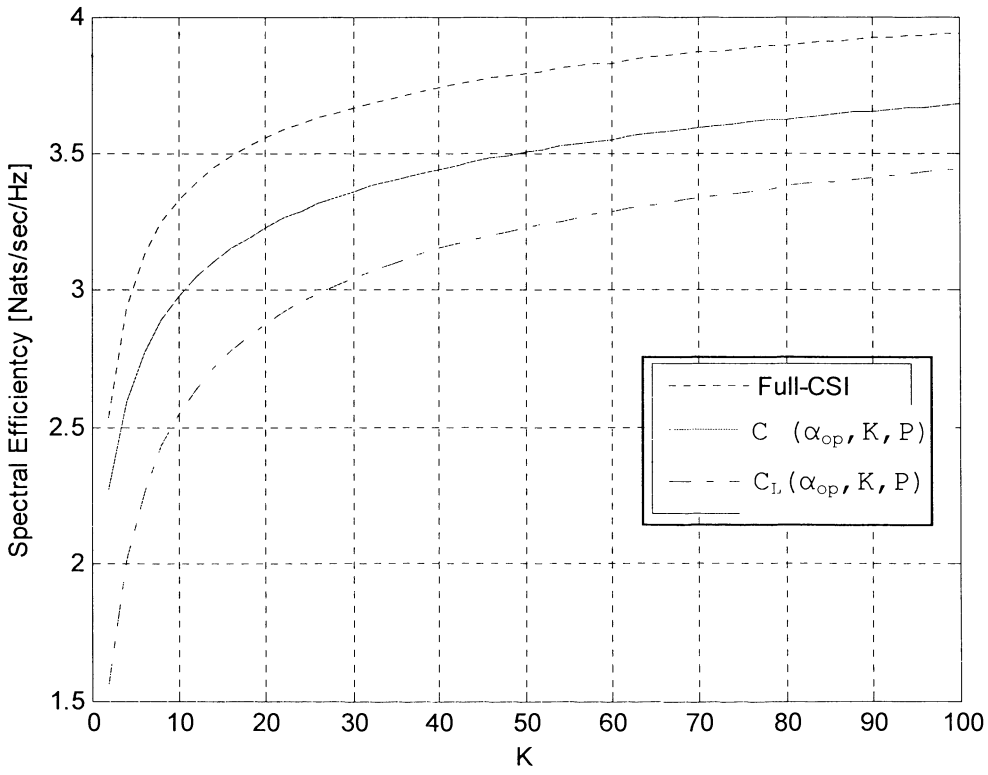


Figure 4.3 The achievable sum-rate with optimal threshold

4.3 Achievable Throughput with Multiple Feedback Thresholds

In this section, a new feedback technique with multiple thresholds is proposed and the effects of the thresholds on the achievable sum-rate of multiuser systems with opportunistic scheduling is analysed.

It is assumed that the system is a single carrier system in which the feedback information are sent via a reliable delay-free uplink channel, and the base station selects only one user to communicate to in each scheduling slot

The proposed multiple thresholds technique requires the mobile users to compare the received SNR with the multiple thresholds and send an appropriate CSI feedback to the base station on every scheduling slot. The computation complexity of the proposed scheme at mobile user is similar to the one-bit feedback scheme. At the base station, the optimum multiple thresholds for feedback can be computed in advance and broadcasted to all users for each corresponding number of users (i.e. a set of optimum feedback thresholds for each number of users). This computation can be performed offline, and thus no additional real-time complexity is required to implement the proposed scheme.

The system with J feedback thresholds $(\alpha_0, \alpha_1, ..., \alpha_{j-1})$ is shown in Figure 4.4.

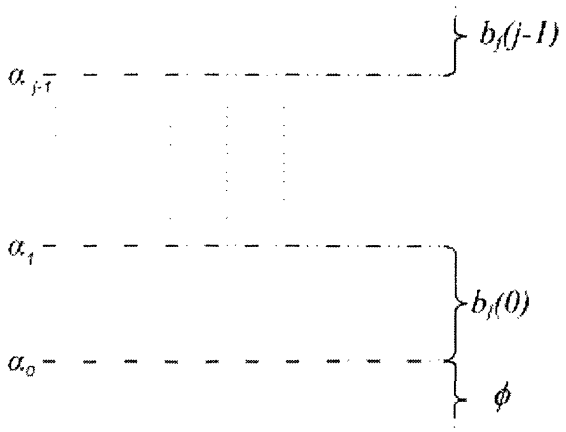


Figure 4.4 System with j feedback thresholds

A unique CSI bit pattern, b_f, is sent when the received SNR is above one of the thresholds shown above. $b_f(0)$ is the CSI bit pattern (one or more bits) transmitted when received SNR is greater than α_0 but less than α_1. $b_f(1)$ is the CSI bit pattern transmitted when received SNR is greater than α_1 but less than α_2 and so on for $b_f(J-1)$. α_j is set at infinity. For example, in a system with three feedback thresholds b_f of user k ($b_{f,k}$) may consists of two-bits and be described as follows

$$b_{f,k} = \begin{cases} \phi & ; \quad \gamma_k < \alpha_0 \\ 01 & ; \alpha_0 \le \gamma_k < \alpha_1 \\ 10 & ; \alpha_1 \le \gamma_k < \alpha_2 \\ 11 & ; \alpha_2 \le \gamma_k \end{cases}. \tag{4.7}$$

where ϕ means that feedback is not sent.

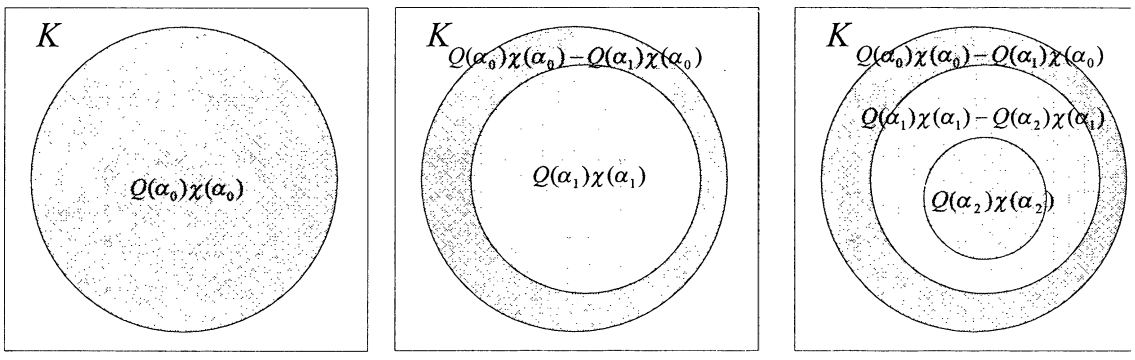
The base station receives the feedback bit(s) sent from all users and it randomly selects one of the users with the highest CSI, i.e. $b_f(j-1)$. If there is no user sending $b_f(j-1)$, it allocates the resource in the scheduling slot t to one of users sending $b_f(j-2)$, $b_f(j-3)$, and so on. If the received SNR of all users is lower than the lowest threshold (α_0), none of users sends CSI and the base station keeps silent for the current scheduling slot. Hence, the base station communicates with at most one user at any time.

To formulate the achievable sum rate of multiuser system with one or more feedback thresholds, let us consider the diagrams in Figure 4.5.

Figure 4.5(a) shows the diagram of multiuser system with single feedback threshold (α_0). Parameters K and P are omitted for simplification. The shaded circle in Figure 4.5(a) represents the sum rate for users whose received SNR is above α_0.

Figure 4.5(b) demonstrates the sum rate of multiuser system with two feedback thresholds (α_0 and α_1). The sum rate of users whose received SNR exceeds α_1 can be represented by the inner circle ($Q(\alpha_1)\cdot\chi(\alpha_1)$). The sum rate of users whose received SNR exceeds α_0 but below α_1 can be represented by the outer donut shape i.e. $Q(\alpha_0)\cdot\chi(\alpha_0)-Q(\alpha_1)\cdot\chi(\alpha_0)$.

Similar procedures as in Figure 4.5(b) can be repeated in Figure 4.5(c) where there are three feedback thresholds (α_0, α_1 and α_2).



(a) (b) (c)

Figure 4.5 Diagram showing the sum rate of multiple thresholds

Based on above explanation of Figure 4.5 (a) to (c), a mathematical expression of the achievable sum rate of multiuser system with J multiple feedback thresholds,($J \geq 2$) can be formulated as follows:

$$C(\overline{\alpha}_J, K, P) = Q(\alpha_{j-1}, K) \cdot \chi(\alpha_{j-1}, P) + \sum_{j=0}^{J-2} (Q(\alpha_j, K) \cdot \chi(\alpha_j, P) - (Q(\alpha_{j+1}, K) \cdot \chi(\alpha_j, P)) \quad (4.8)$$

where $\overline{\alpha}_J$ is a $1xJ$ dimension vector whose elements are feedback thresholds, $\overline{\alpha}_J = \{\alpha_0, ..., \alpha_{J-1}\}$ ..

The first term in the right hand side of Equation 4.8 may be viewed as the sum rate of the inner circle in Figure 4.5 and the sigma term may be viewed as the summation of all donut-shape sum rates.

The lower bound of Equation 4.8 can be obtained by replacing $\chi(\alpha_{j-1}, P)$ with $\chi_L(\alpha_{j-1}, P)$ from Equation 4.5.

The set of optimal thresholds, $\acute{\alpha}$, for each value of K can be calculated from a J-dimensions search to satisfy the following criteria:

$$\alpha'(K) = \arg\max_{\overline{\alpha}} C(\overline{\alpha}, K, P) \quad (4.9)$$

4.4 Numerical Results

In Figure 4.6, the achievable sum-rate of the proposed scheme in Equation 4.8 and compared with the simulation results from MATLAB-based Monte Carlo simulation when two feedback thresholds are employed (i.e. J = 2). The channel model is Rayleigh block fading, the feedback channel is assumed error free and delay free. The average of receives SNR is 10 dB.

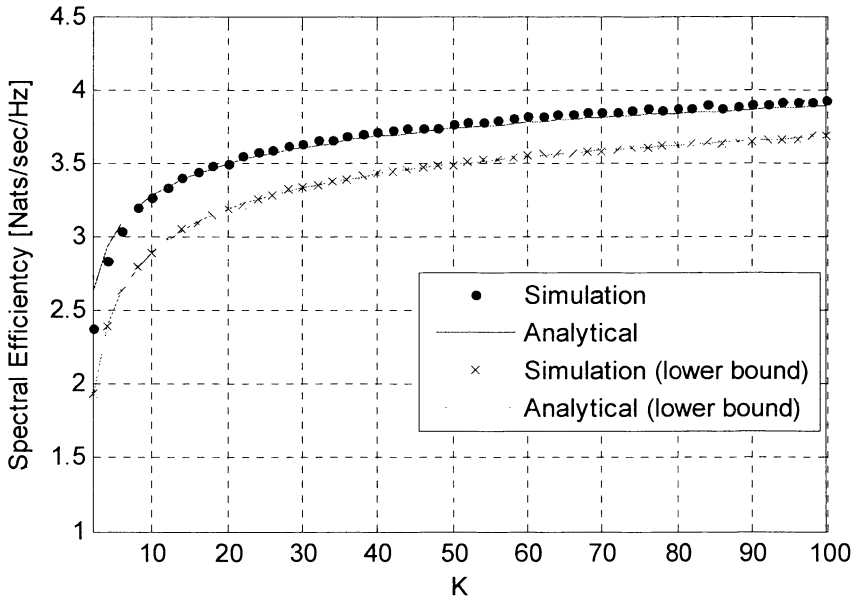


Figure 4.6 Validation of the proposed closed-form expression

Figure 4.6 shows that the result from the proposed analytical equation matches the simulation well for both the achievable sum rate and its lower bound. More specifically, the difference between simulation and analytical results is less than 0.1%. The difference diminishes when the number of users become larger. The optimal threshold values are obtained by extensive search on Equation 4.9. It is not mathematically feasible to calculate the optimal feedback thresholds even for a single threshold system as stated in [21] [22].

In Figure 4.7, the achievable sum-rate and its lower bound are plotted versus the number of feedback thresholds when the number of users K =200., $R_{Proposed}$ represents the achievable sum-rate of the proposed feedback scheme with multiple thresholds, and $R_{L\ Proposed}$ represents its lower bound.

The achievable sum rate normalised by the sum rate of full-CSI when the number of feedback thresholds J = 1 to 8 is summarised in Table 4.1.

Table 4.1 Achievable sum rate for $J = 1$ to 8 compared with full-CSI

Number of thresholds, J	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Achievable sum rate normalised with full CSI	94.10%	98.15%	99.11%	99.48%	99.66%	99.76%	99.81%	99.84%

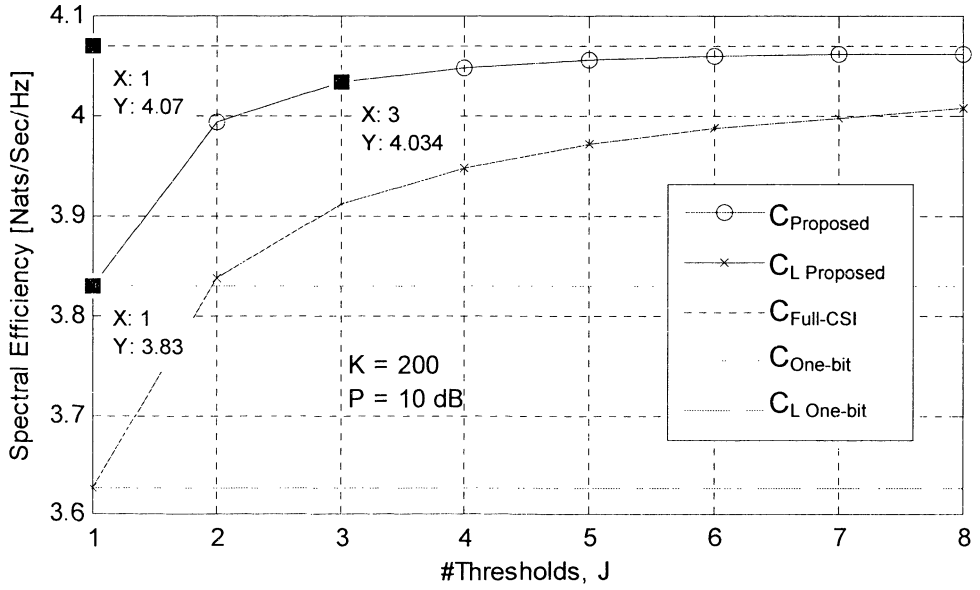


Figure 4.7 Spectral efficiency versus number of thresholds

Figure 4.7 indicates that the achievable spectral efficiency and its lower bound increase as a function of the number of feedback thresholds can be observed. The system with single feedback threshold can obtain the achievable sum rate of 94.10% compared with full-CSI. For the proposed feedback scheme with two thresholds, the achievable sum rate is 98.15% compared with full-CSI. For three feedback thresholds, the achievable sum-rate is as high as 99.11%. Given that three feedback thresholds require two bits to represent all possible combinations of b_f, the 0.96 % improvement is considered to be trivial compared with the amount of feedback bit increased (from one bit to two bits). Hence, two feedback thresholds are selected as an optimal configuration in terms of the balance between achievable sum-rate and feedback load.

4.4.1 Performance with Two Feedback Thresholds

In the previous subsection, the performance of the proposed feedback scheme with three feedback thresholds was studied. Although it can achieve the sum-rate performance very close to the full-CSI, an additional bit in CSI message is required when compared with the single feedback threshold case. In this subsection, the feedback technique employing two feedback thresholds using only one bit CSI is proposed and analysed.

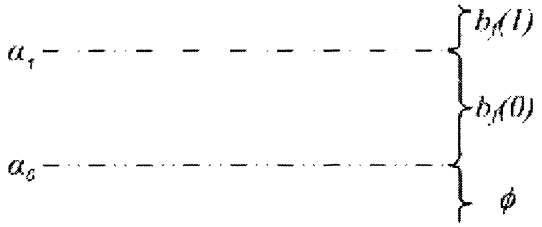


Figure 4.8 System with two feedback thresholds

The system with two feedback thresholds is shown in Figure 4.8. The feedback threshold can then be represented by:

$$b_{f,k} = \begin{cases} \phi & ; \quad \gamma_k < \alpha_0 \\ 0 & ; \alpha_0 \le \gamma_k < \alpha_1 \\ 1 & ; \alpha_1 \le \gamma_k \end{cases}. \tag{4.10}$$

where ϕ means no feedback is sent.

The expression in Equation 4.10 indicates that one-bit is sufficient to represent two CSI bit patterns when two feedback thresholds are employed.

The achievable sum rate of the system with two optimally selected feedback thresholds is plotted in Figure 4.9.

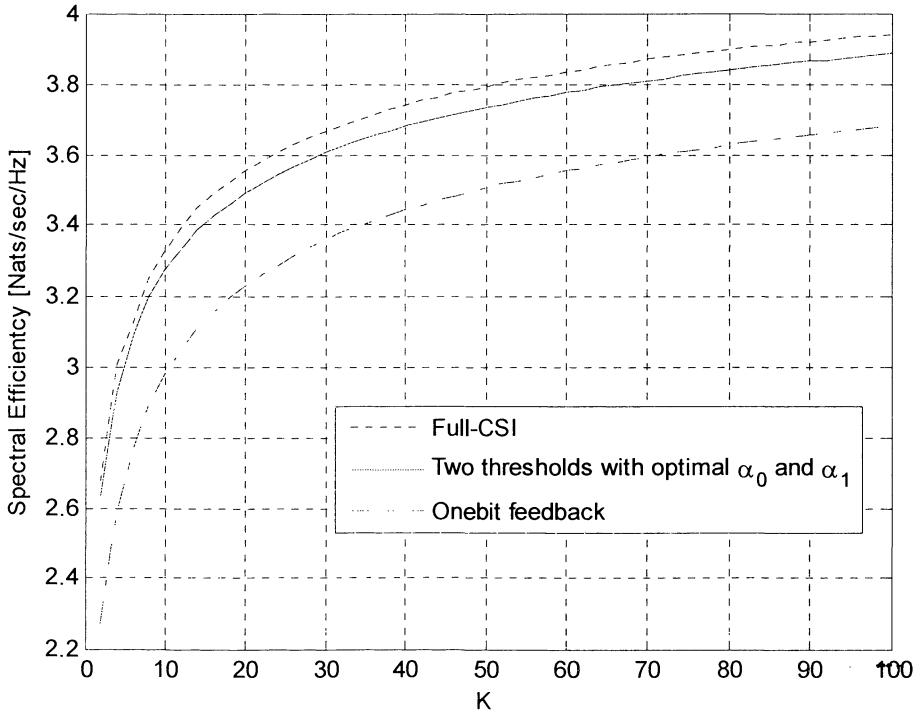


Figure 4.9 Achievable sum-rate capacity with two feedback thresholds, P = 10dB

Table 4.2 Achievable sum rate for J = 1 and 2 for K = 2 to 100

Number of users, *K*	**2**	**5**	**10**	**20**	**50**	**100**	**Average**
Achievable sum rate normalised with full CSI (J =2)	98.46%	98.00%	98.36%	98.26%	98.42%	98.62%	98.36%
Achievable sum rate normalised with full CSI (J =1)	85.04%	87.38%	89.52%	90.87%	92.42%	93.43%	89.78%
% Improvement comparison (J = 1 vs J =2)	13.42%	10.61%	8.83%	7.38%	5.99%	5.18%	8.57%

It can be observed from Figure 4.9 and Table 4.2 that the proposed feedback scheme with two thresholds can achieve 8.57% improvement on average in sum rate when compared with the one-bit feedback scheme with single threshold. The values of optimal dual thresholds to achieve this 8.57% improvement can be calculated offline, similar to single feedback threshold case. Hence, there is no additional feedback load over air interface. Given that the proposed two feedback thresholds scheme uses the same amount of feedback as one bit feedback scheme with single threshold, the proposed scheme provides more efficient utilisation of feedback resources to obtain higher sum-rate compared with having single feedback threshold.

In the following subsection, the effect of the varying the feedback thresholds (α_0 and α_1) on the sum-rate performance is evaluated:

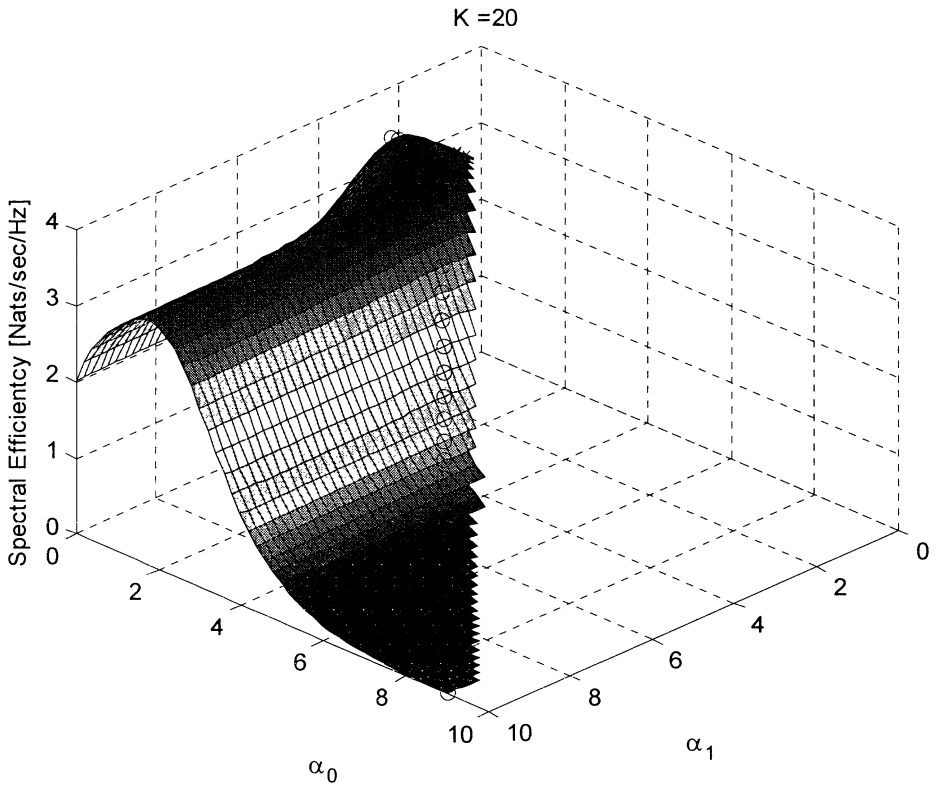


Figure 4.10 Achievable sum-rate as function of α_0 and α_1 when K =20

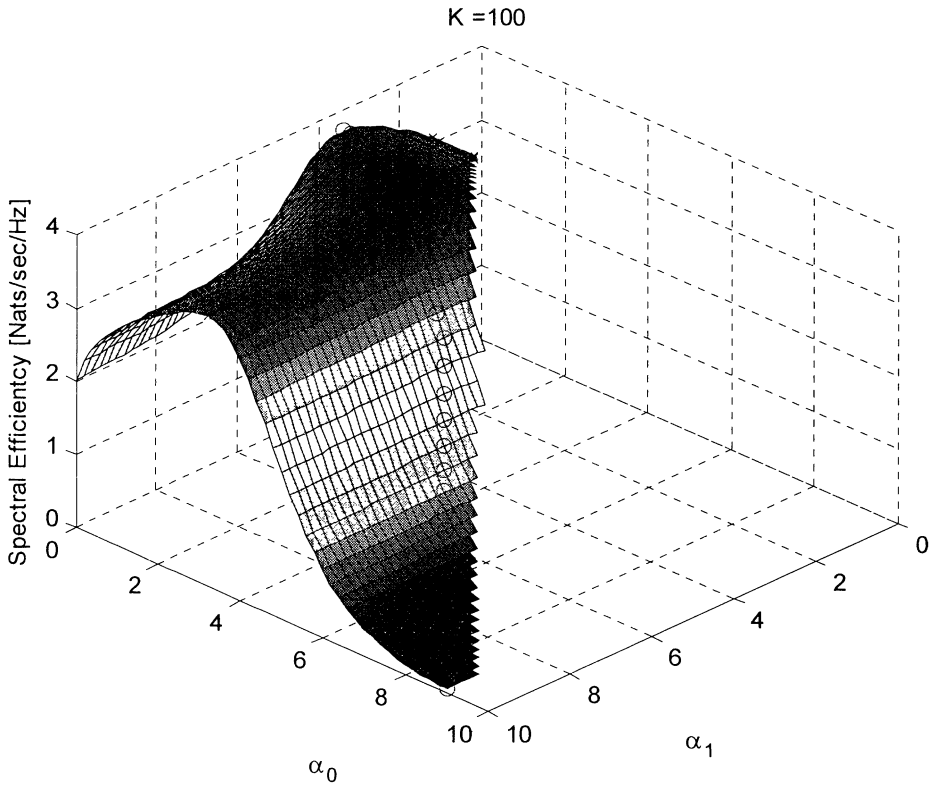


Figure 4.11 Achievable sum-rate as function of α_0 and α_1 when K =100

Figure 4.10 and 4.11 show the achievable sum-rate of the multiuser system with two feedback thresholds for K = 20, and 100 respectively. The local optimal values of α_1 for each α_0 are highlighted by circles. Optimal values of α_0 and α_1 represent the feedback thresholds at which spectral efficiency is a maximum (the intersection of the local optimal lines).

It can be observed from Figure 4.10 and 4.11 that the performance of the proposed feedback scheme with two thresholds is sensitive to the choice of the value of feedback thresholds.

Figure 4.12 shows the optimal α_0 and α_1 for K = 2 to 100 in comparison with optimal α for single threshold system (from Figure 4.2).

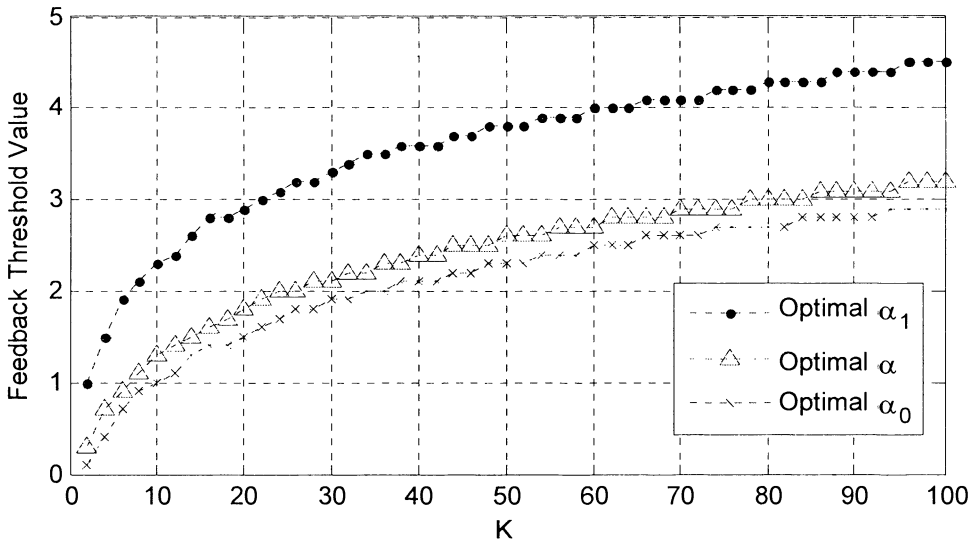


Figure 4.12 Optimal α_0 and α_1 versus number of users K

Figure 4.12 demonstrates that the values of optimal α_1 in the two feedback thresholds system are relatively large compared with the values of α_0 in the single feedback threshold system. The larger value of optimal α_1 is more desirable as it increases the possibility that the base station would select a user with a high received SNR in each scheduling slot.

It can be observed from Figure 4.12 that the optimal values of α_0 in the system with two feedback thresholds are lower than the optimal threshold in the system with single feedback threshold (optimal α). Lower value of α_0 is beneficial to the system as the feedback outage would be lower. Figure 4.13 compares the feedback outage of system employing one feedback and two feedback thresholds respectively.

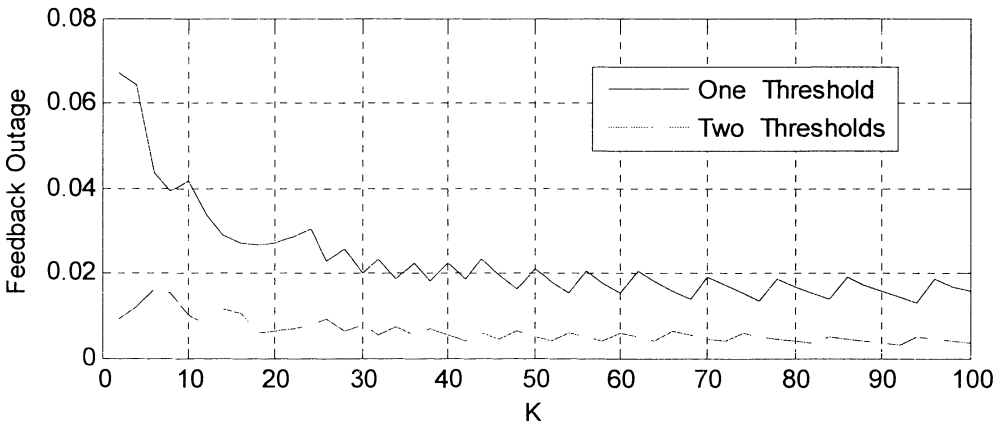


Figure 4.13 Comparison of feedback outage

It can be observed from Figure 4.13 that the feedback outage in system with two feedback thresholds is well below half of the outage of single threshold system. This is because the optimal values of α_0 in the system with two feedback thresholds are lower than the optimal thresholds in the system with single feedback threshold (optimal α).Accordingly, the possibility that all users would have received SNR lower than α_0 decreases is lower with two feedback thresholds.

4.5 Analysis of Suboptimal Feedback Thresholds

In this section, the sub-optimal setting of feedback thresholds is investigated to further evaluate the performance of multiuser system with two feedback thresholds. The analysis starts with determining the suboptimal feedback thresholds for single feedback system. Then, the suboptimal thresholds in a multiuser system with multiple feedback thresholds are analysed.

4.5.1 Suboptimal Threshold for System with Single Feedback Threshold

As observed in Chapter 3, the average number of users sending feedback, K_{FB}, when the feedback thresholds are optimally selected is nearly constant. Based on this observation, a hypothesis that optimal feedback threshold would keep the average number of users sending feedback at constant is investigated as below..

Hypothesis: *for any k there is* $\alpha_{op}(K)$ *that* $K_{FB}(K) \approx constant.$ (4.11)

when K is the number of users.

In a Rayleigh fading channel, the average number of users having received SNR that exceeds the feedback threshold α can be express as:

$$K_{FB}(K) = K \cdot exp(-\alpha) \tag{4.12}$$

Assuming $K_{FB}(K)$ is a constant for any K, the index K can be omitted ($K_{FB}(K) = K_{FB}$). The optimal (suboptimal) feedback threshold can be calculated using:

$$\alpha_{op} = log(K/K_{FB}) \tag{4.13}$$

.The problem is to find a constant K_{FB} that is suitable for all K. To evaluate the optimal value of K_{FB}, the spectral efficiency (Equation 4.1) where the value of α is calculated from Equation 4.13 is plotted in Figure 4.14 as a function of K and K_{FB}.

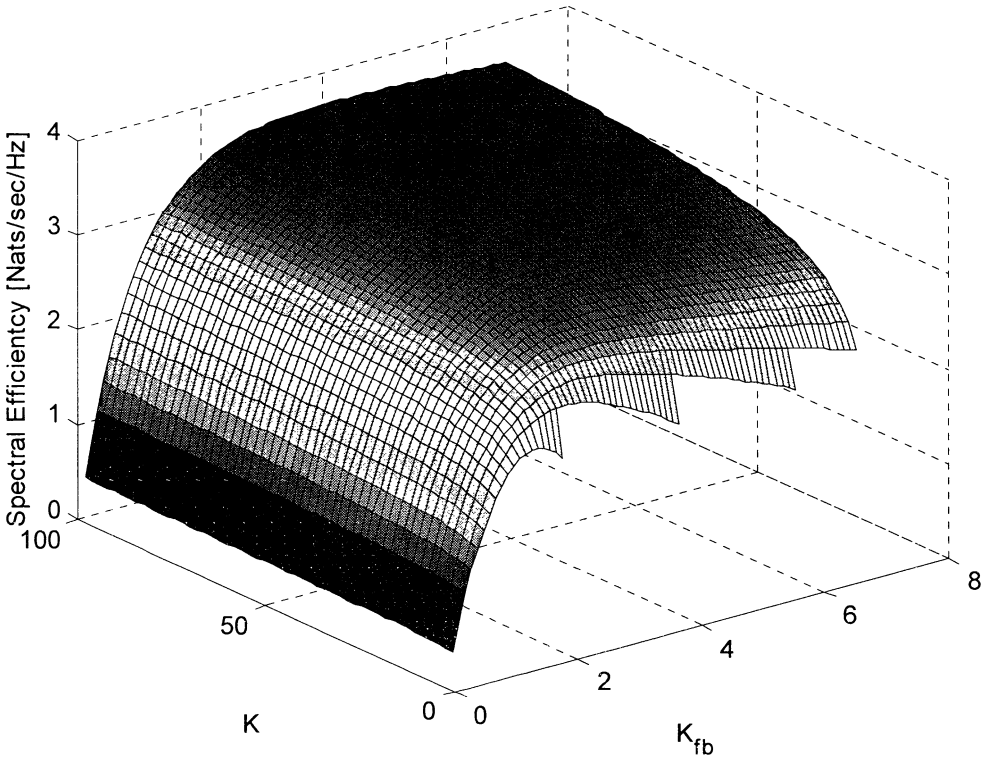


Figure 4.14 Spectral efficiency when threshold is calculated from Equation 4.13.

Figure 4.14 illustrates that the spectral efficiency performance varies when the constant K_{FB} changes.

To select a suitable value for K_{FB}, the values of K_{FB} that can provide the highest sum rate for a given value of K is plotted in Figure 4.15.

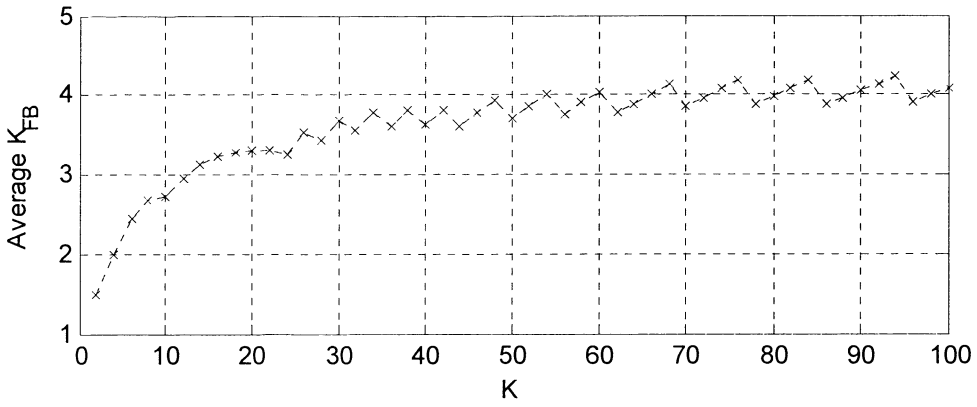


Figure 4.15 Average number of users sending feedback at different K.

Figure 4.15 shows that the average number of users who send feedback from a user range of 2 to 100 users is 3.6. This number will be inserted into Equation 4.13 to generate suboptimal thresholds.

For comparison, the sum rate performance of the system with optimal feedback thresholds as derived in [22]:

$$\alpha_{op} = log\ K - log\ log\ K \tag{4.14}$$

will be compared with suboptimal threshold computed from Equation 4.13 with K_{FB} = 3.6 while sum rate with optimally selected thresholds as given in Figure 4.2 is plotted as a reference.

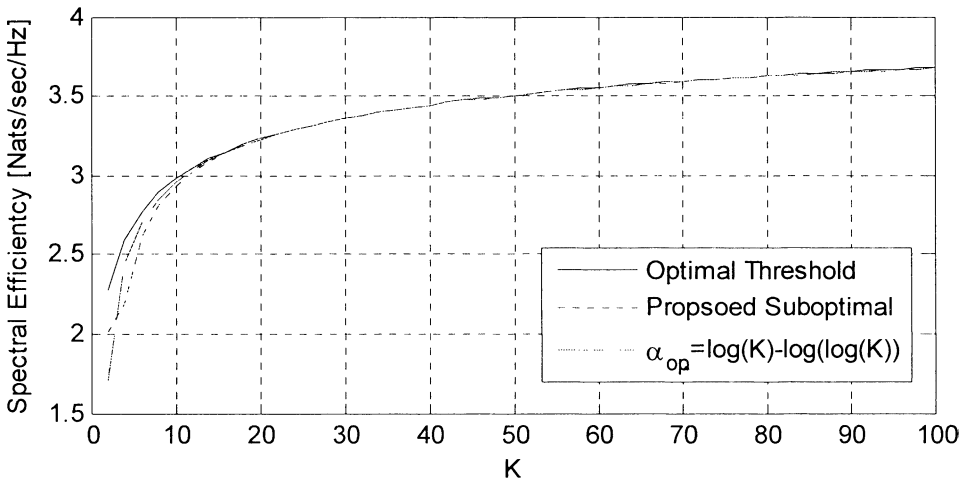


Figure 4.16 Spectral efficiency of system with suboptimal feedback thresholds

Figure 4.16 shows that the achievable sum rate obtained from suboptimal feedback thresholds of the proposed scheme (Equation 4.13) matches the sum rate of the suboptimal thresholds from Equation 4.14 well, especially when the number of users becomes large. The result validates the hypothesis in Equation 4.11. This leads to a conclusion that the suboptimal feedback threshold can be set as a function of the mean value of number of users sending feedback. This result will be beneficial for designing a feedback channel particularly contention-based feedback techniques.

4.5.2 Suboptimal Thresholds for System with Two Feedback Thresholds

In this section, the analysis for suboptimal feedback thresholds is extended to a multiuser system with two feedback thresholds. Note that the method to determine optimal feedback thresholds in [22] is not applicable in a multiple threshold scenario.

In the previous section, the values of optimal α_0 and α_1 are plotted in Figure 4.12. To evaluate the relationship between these thresholds, the ratio of the two thresholds is plotted in Figure 4.17 below. Figure 4.17 illustrates the ratio (in dB) between the optimal values of α_o and the optimal value of α_1 for $K = 2$ to 100. dB.

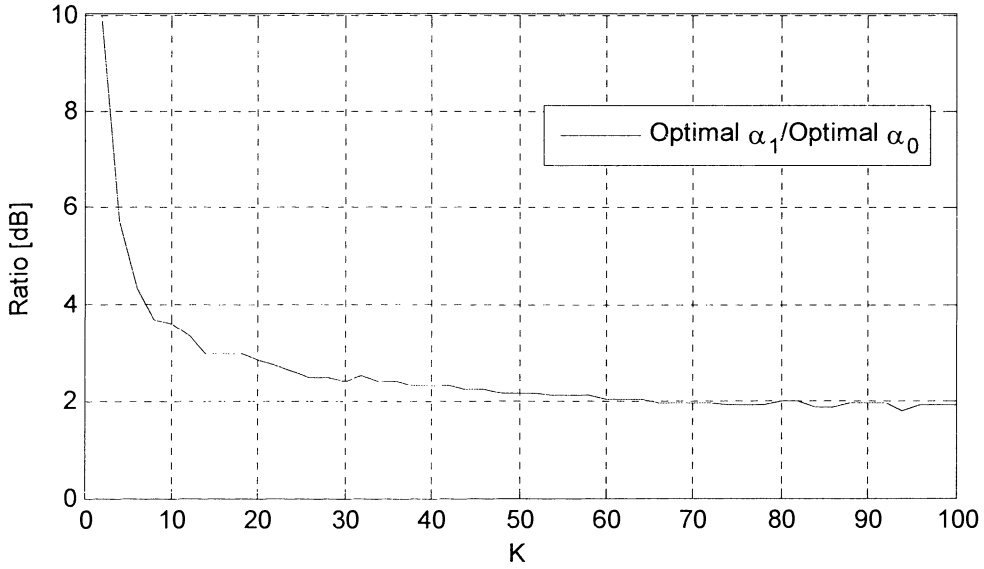


Figure 4.17 Ratio of optimal α_o and optimal α_1 for each number of users K.

Figure 4.17 indicates that the ratio between the feedback thresholds converges to 2 dB when the number of users becomes larger. This result suggests that the sub-optimal solution to determine the value of feedback thresholds exists. Based on this observation, the suboptimal values of feedback thresholds α_{o_s} and α_{1_s} can be expressed as:

$$\alpha_{1\_s} = \alpha_{o\_s} + 2\text{dB} \qquad \text{[in dB]} \qquad (4.15)$$

or

$$\alpha_{1\_s} = 10^{0.2} \cdot \alpha_{o\_s} \qquad (4.16)$$

Figure 4.18 shows the achievable sum-rate of the proposed feedback scheme with suboptimal feedback thresholds which is now reduced to a function of only. The value α_{o_s} can be determined based on criteria in Equation 4.9 which is now reduced to a function of α_{o_s} only. The local suboptimal values for each number of users are highlighted in blue circles.

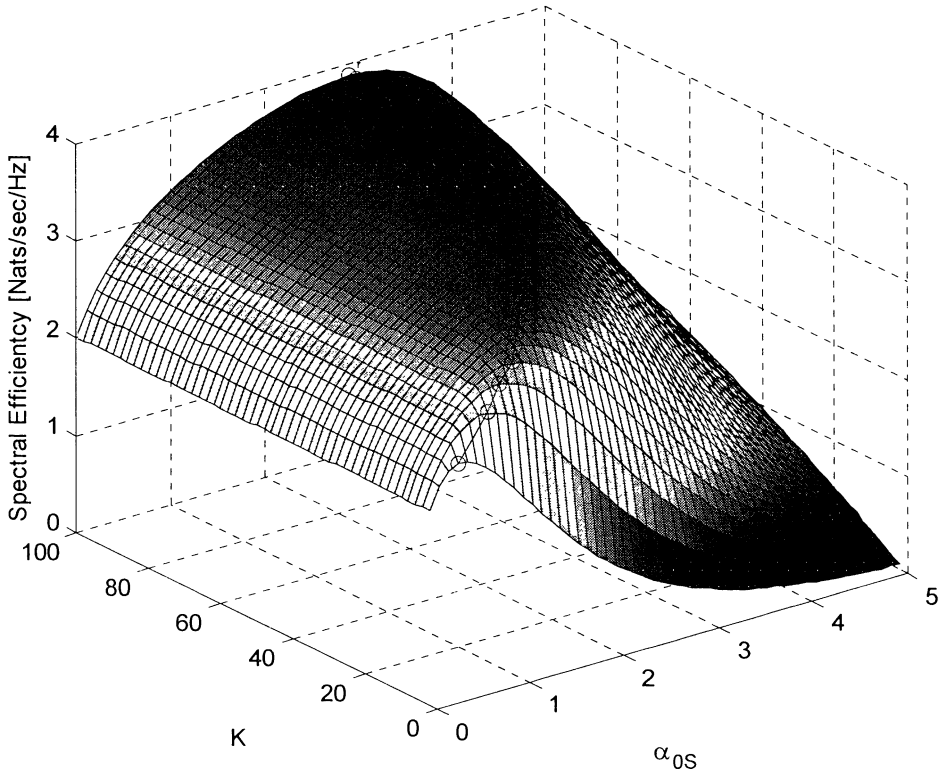


Figure 4.18 Achievable sum-rate of system with two feedback thresholds

The achievable sum-rate using the suboptimal feedback thresholds is plotted in Figure 4.19.

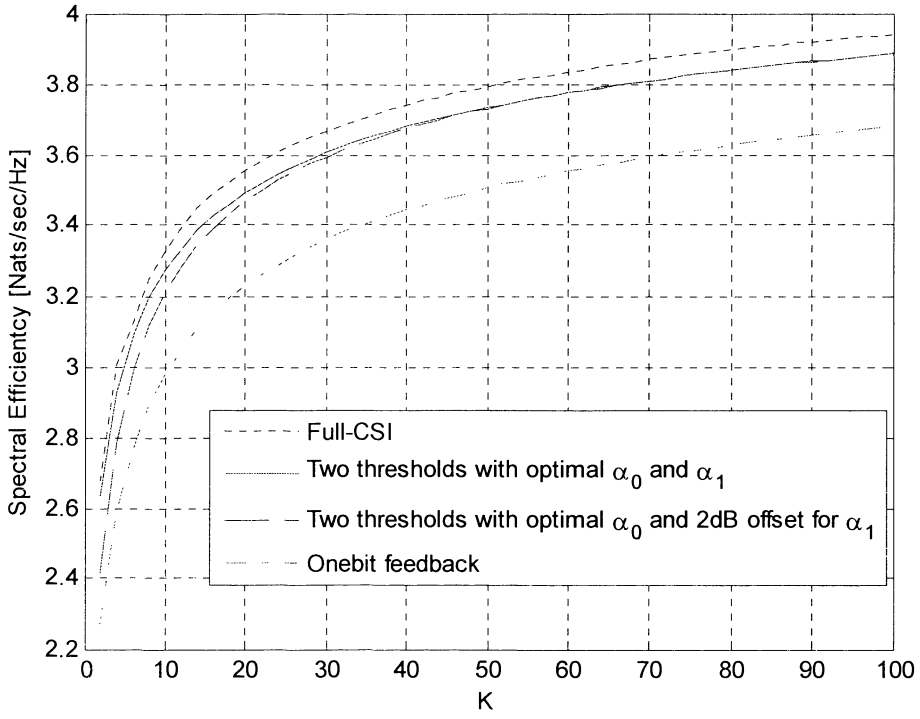


Figure 4.19 Comparison of achievable sum-rate

Figure 4.19 shows the sum rate of the proposed feedback technique with suboptimal thresholds (Equation 4.16) in comparison to full-CSI, two feedback thresholds with optimally selected thresholds and one-bit feedback with single optimally selected threshold.

The result in Figure 4.19 indicates that the sum rate of achieved by using suboptimal feedback thresholds from Equation 4.16 is close to the sum rate of two thresholds technique with optimally selected thresholds. There is a little performance gap when the number of users is low. The gap diminishes when the number of users grows.

Figure 4.20 shows the spectral efficiency of the proposed feedback technique with two thresholds and that of the one-bit feedback technique normalised by the achievable spectral efficiency of the system with full CSI.

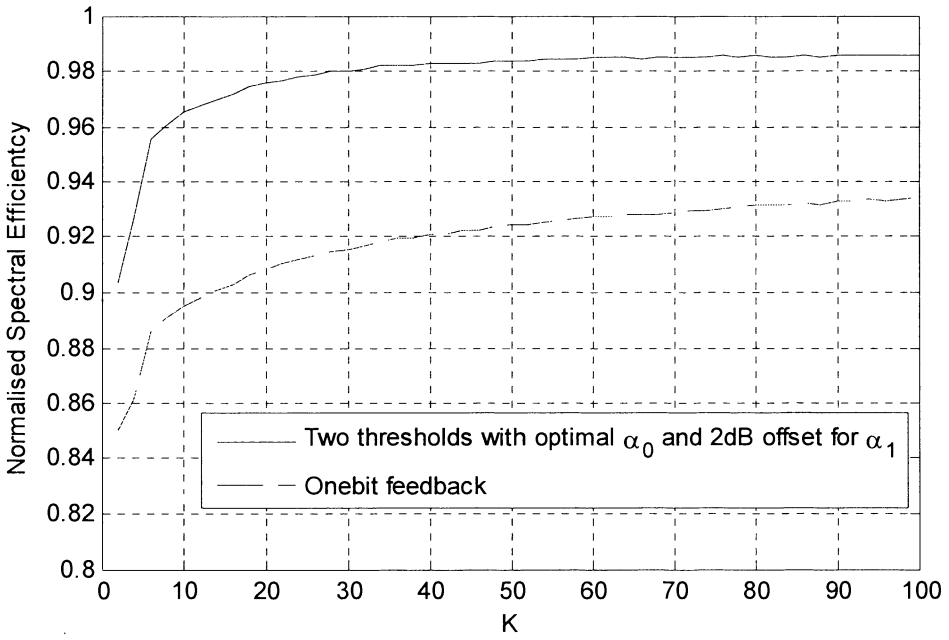


Figure 4.20 Normalised spectral efficiency

From Figure 4.20, the proposed technique improves the achievable spectral efficiency approximately 5% without any need for additional feedback bit in CSI message.

It can be concluded that the use of multiple feedback thresholds results in a more efficient utilisation of feedback resource. The additional complexity is negligible compared with the one-bit technique.

4.6 Summary

In this chapter, the achievable sum rate of multiuser system with multiple feedback thresholds is formulated. Numerical result showed that one bit feedback with two feedback thresholds can improve the achievable sum rate by around 8.5% compared with one bit feedback with single feedback threshold. The formulation is used to evaluate the suboptimal feedback thresholds in system with two thresholds. The suboptimal α_1 can be generated by applying 2 dB to α_0.

Chapter 5

Automatic Feedback Threshold Setting and Hybrid Scheduling Technique

This chapter consists of two parts: In the first part, a new technique in which feedback threshold is automatically adjusted at each mobile user in single carrier multiuser system is proposed in Section 5.1. Simulation results showing the performance of the proposed automatic feedback threshold setting are discussed in Section 5.2.

In the second part of this chapter, problems associated with scheduling delays are addressed in Section 5.3 where scheduling techniques presented in the literature to mitigate the delay problems are overviewed. In Section 5.4, a hybrid scheduling technique to minimise the scheduling delays in heavily loaded single carrier network is presented. Simulation results showing the performance of the proposed hybrid scheduling are discussed in Section 5.4. Section 5.5 further demonstrates the performance of the proposed hybrid technique over a well-known delay-aware scheduling algorithm in heavily loaded scenario. Section 5.6 concludes this chapter.

5.1 Automatic Feedback Threshold Setting

As discussed in the previous chapters, allowing users whose received SNR is above a predefined threshold can significantly reduce the amount of feedback messages. This section presents a technique for the users to automatically adjust the feedback threshold.

5.1.1 Overview of Feedback Threshold Adjustment Techniques

Gesbert et al suggested that the feedback threshold can be calculated as expressed below as a function of feedback outage and the number of users [20]:

$$\gamma_{th} = -\log(1 - P_{ot}^{1/K}) \tag{5.1}$$

where log(·) is a natural logarithm and P_{ot} is the probability of feedback outage. The feedback outage probability can be pre-set. The base station updates the feedback threshold when the number of users in the coverage area changes and broadcasts the information to all users to update the threshold.

In [23], Kim et al proposed a feedback reduction technique in which the feedback threshold is automatically adjusted to meet a targeted number of users sending feedback as discussed in Chapter 3. The automatic updating was performed in a centralised manner at the base station which has to constantly update the feedback threshold in every scheduling slot. This implies that additional radio resources will be required to implement this scheme

5.1.2 Proposed Automatic Feedback Threshold Adjustment Technique

A new feedback reduction technique is proposed in this section to eliminate the need to frequently update the value of feedback threshold in every scheduling slot. The proposed feedback technique schedules a user in each scheduling slot based on the proportional fair policy using the relative SNR. It allows only those users having relative SNR above the feedback threshold to send CSI. The difference of the proposed technique with the technique in [20] is that the proposed scheme does not require the base station to broadcast the value of feedback threshold when the number of users in the system changes. In other words, the proposed feedback reduction scheme updates the feedback threshold in a distributed manner where all users locally calculate the feedback threshold based on the following equation:

$$\gamma_{th}(k,t+1) = \begin{cases} \gamma_{th}(k,t)/\Delta\gamma_{dw} & when\ user\ k^{th}\ is\ scheduled\ in\ t^{th}\ slot \\ \gamma_{th}(k,t)+\Delta\gamma_{up} & else \end{cases} \tag{5.2}$$

and

$$\gamma_{th}(k,t+1) < \gamma_{th,\max} \tag{5.3}$$

where feedback threshold increasing step $\Delta\gamma_{up}$ and feedback threshold decreasing factor $\Delta\gamma_{dw}$ are constants representing feedback threshold adjustment factor.

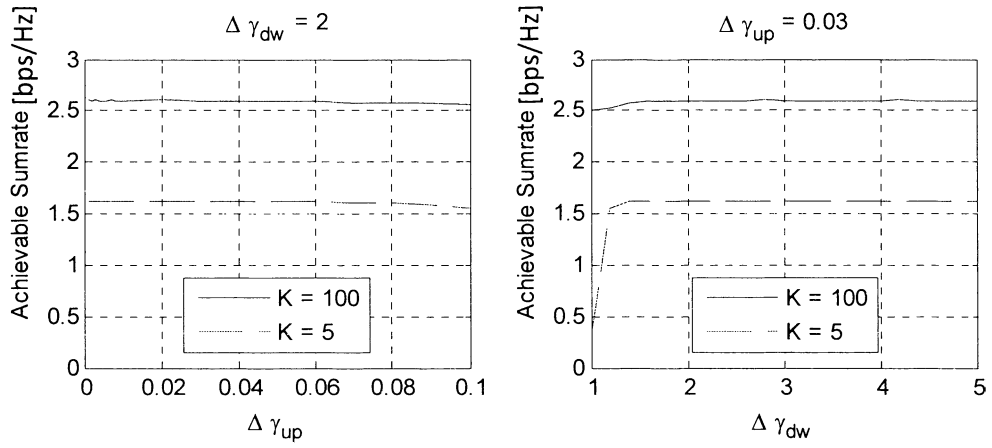
Parameter $\gamma_{th\ max}$ is defined as the maximum value of feedback threshold (set as 3.5 or $10 \cdot \log_{10}(3.5) = 5.4$ dB). The optimal value of $\gamma_{th\ max}$ is determined by Equation 5.1 where the maximum number of users $K = 100$ and $P_{ot} = 0.03$ i.e. $\gamma_{th} = -\log(1-0.03^{1/100}) = 3.368$ and rounded up to 3.5. $P_{ot} = 0.03$ was chosen because it was observed that the sum rate starts to noticeably drop when P_{ot} is greater than 0.03.

From Equation 5.2, the proposed feedback reduction technique gradually increases the value of the feedback threshold as long as the user is not scheduled for access to radio resource in the t^{th} scheduling slot.

The basic idea behind the proposed technique is that each user slowly increases its feedback threshold in each scheduling slot aiming to minimise the possibility to send feedback messages. The feedback threshold will decrease when the user is scheduled to ensure that the feedback threshold is not set too high. It is critical that the feedback threshold is gradually increased by a relatively small step otherwise the possibility that users do not send feedback (feedback outage) would be dramatically high. Thus, the feedback threshold must not grow at an excessive rate. When the number of users increases it would be more likely that the feedback threshold of each user keeps growing as the probability that a user will be scheduled becomes proportionally less. As a result, the value of feedback threshold is larger when the number of users increases which matches the requirement of a larger value of feedback threshold for a higher number of users.

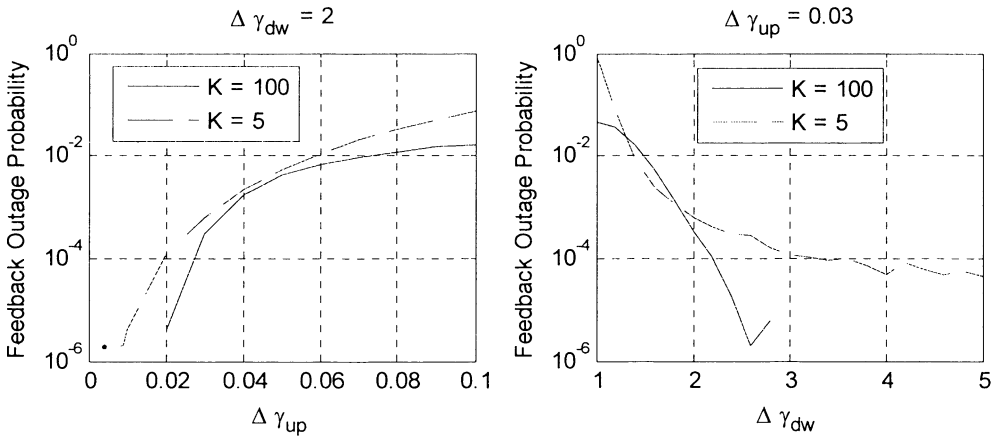
The performance of the proposed feedback reduction technique relies on the setting of parameters $\Delta\gamma_{dw}$ and $\Delta\gamma_{up}$. Therefore, these parameters need to be carefully selected. The simulation results necessary for optimising these parameters are given in Figure 5.1, Figure 5.2 and Figure 5.3. Figure 5.1 to 5.3 show the performance of the proposed feedback reduction techniques in terms of achievable sum-rate, feedback outage probability and feedback load (the number of users sending feedback, K_{fb}) for different value of $\Delta\gamma_{dw}$ and $\Delta\gamma_{up}$ as labelled in the corresponding figure.

For Figure 5.1(a), 5.2(a), and 5.3 (a), the parameter $\Delta\gamma_{dw}$ is set at 2 based on initial simulations indicating that the changes in $\Delta\gamma_{up}$ with the parameters within the range $\Delta\gamma_{dw} \in [1, 5]$ had the same tend. Similar approach was used to set $\Delta\gamma_{up}$ in Figure 5.1(b), 5.2(b), and 5.3 (b). It will be shown that the proposed technique can provide satisfactory performance over a wide range of $\Delta\gamma_{up}$ and $\Delta\gamma_{dw}$.



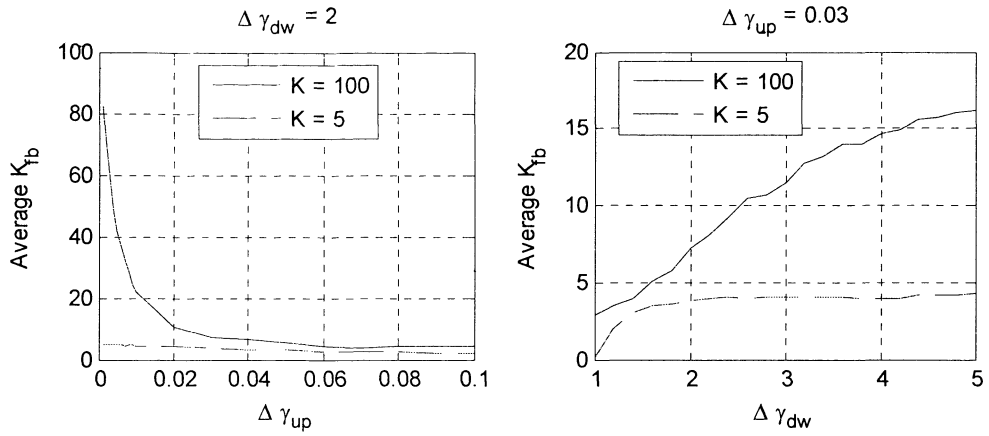
(a) (b)

Figure 5.1 Achievable sum rate of the proposed feedback reduction technique



(a) (b)

Figure 5.2 Feedback outage probability of the proposed feedback reduction technique



(a) (b)

Figure 5.3 Average number of users sending feedback of the proposed technique

The selection of these parameters is analysed in the following subsections:

5.1.3 Selection of parameter $\Delta\gamma_{up}$

Figure 5.1(a) shows that the selection of $\Delta\gamma_{up}$ has negligible effect on the sum-rate performance within the range of $\Delta\gamma_{up} \in [0.001, 0.1]$. However, the sum-rate starts to noticeably drop when the value of $\Delta\gamma_{up}$ exceeds 0.1 (not shown in the figure). This is because the feedback threshold increases rapidly and most users may not have received SNR above the feedback threshold. Hence, $\Delta\gamma_{up}$ must not be greater than 0.1.

Figure 5.2(a) indicates that feedback outage probability increases as the value of the $\Delta\gamma_{up}$ increases. As mentioned earlier, a high value of $\Delta\gamma_{up}$ would make the γ_{th} reaches $\gamma_{th\ max}$ and thus very few or none of users would have their relative SNR above the feedback threshold in most scheduling slots. As a result, the feedback outage probability increases substantially with an increase of $\Delta\gamma_{up}$ as depicted in Figure 5.2.

Figure 5.3(a) shows that feedback load decreases when a larger value of $\Delta\gamma_{up}$ is chosen. This result demonstrates that the proposed feedback reduction technique is capable of reducing feedback load effectively. However, as previously discussed a

large $\Delta\gamma_{up}$ results in performance loss due to high feedback outage probability, thus $\Delta\gamma_{up}$ must be chosen so that the feedback outage target is met.

As it is observed that the sum rate starts to noticeably drop when feedback outage exceeds 0.03 (3%), the system should be configured such that the feedback outage probability is below 0.03 . Therefore, a largest possible value of $\Delta\gamma_{up}$ that can maintain the outage probability below 0.03 is chosen. Hence, the suitable value of $\Delta\gamma_{up}$ is set at 0.06.

5.1.4 Selection of parameter $\Delta\gamma_{dw}$

Figure 5.1(b) shows that the performance in terms of the achievable sum rate reduces when $\Delta\gamma_{dw}$ is set at a low value. A small value of $\Delta\gamma_{dw}$ is detrimental to the sum rate performance because the feedback threshold would reduce in a slow rate. Thus, the chance that the value of feedback threshold reaches the maximum would be high. This effect leads to a higher feedback outage probability as shown in Figure 5.2(b).

Figure 5.3(b) shows that adjusting the value of $\Delta\gamma_{dw}$ has a significant impact on the feedback load. Although an inappropriately small $\Delta\gamma_{dw}$ would reduce the feedback load, it introduces substantial performance loss due to feedback outage as discussed earlier. In contrast, if too large $\Delta\gamma_{dw}$ was selected, it leads to a situation where feedback threshold is reduced at a rapid rate. As a result, the value of feedback threshold is small and a large number of users would send feedback.

From the feedback load performance perspective, a smaller $\Delta\gamma_{dw}$ would be more desirable provided that the feedback outage probability target of 0.03 is not exceeded. Therefore, a value for $\Delta\gamma_{dw} = 1.5$ is selected.

5.2 Performance of Proposed Feedback Reduction Technique

In this section, the performance of the proposed feedback reduction technique is compared with the performance of the B-bits selective feedback reduction

technique where the value of feedback threshold is set as per Equation 5.1 with P_{ot} = 0.03. The parameters of the simulation are summarised as follows:

Table 5.1 Simulation Parameters

Description	Selective Feedback	Automatic threshold setting
Average SNR	0 dB	0 dB
Users characteristics	Homogeneous (same average SNR)	Homogeneous (same average SNR)
$\Delta\gamma_{up}$	N/A	0.06
$\Delta\gamma_{dw}$	N/A	1.5
$\Delta\gamma_{th\ max}$	N/A	3.5
P_{ot}	0.03 (3%)	0.03 (3%)
Scheduling policy	Simplified PF	Simplified PF
Optimized Threshold	Base station to calculate as per Equation 5.1	Self adjusting (Equation 5.2)
Scheduling slot	1ms	1ms
Channel model	Uncorrelated Rayleigh	Uncorrelated Rayleigh

There are many definitions of fairness. In this work the fairness parameter as introduced in [29] is selected.

As the users are homogeneous, the long-term fairness for the access probability for the multiuser system can be defined as [29]:

$$fr = -\sum_{k=1}^{K} v_k \frac{\log(v_k)}{\log(K)} \tag{5.4}$$

where v_k is the proportion of resources allocated to user k (i.e. the proportion of the time slots allocated to user k) and K is the number of users.

If all users are allocated the same resource, then f_r = 1 i.e. the system is fair.

5.2.1 Sum Rate Performance

The performance of the proposed feedback reduction technique in terms of achievable sum rate is plotted in Figure 5.4:

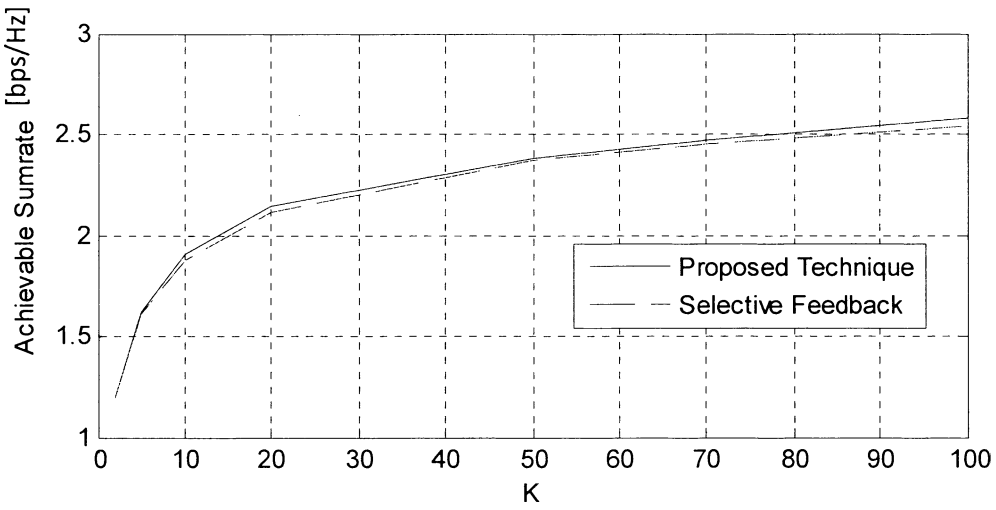


Figure 5.4 Achievable sum rate of the proposed feedback reduction technqiue

Figure 5.4 shows that the proposed feedback algorithm can achieve similar performance to that of original selective feedback with optimal thresholds selected as per Equation 5.1. More specifically, the proposed algorithm can achieve a slightly higher sum rate when compared with the selective feedback technique. This is because the proposed feedback technique with the chosen parameters $\Delta\gamma_{dw}$ and $\Delta\gamma_{up}$ allows more users to send feedback, resulting in lower feedback outage while introducing slightly more feedback load as demonstrated in Section 5.2.2.

5.2.2 Feedback Load and Feedback Outage Performance

Figure 5.5 and Figure 5.6 show the feedback outage probability and the average number of users sending feedback for the proposed feedback reduction technique in comparison with the selective feedback technique for various numbers of users.

From Figure 5.5, it can be observed that the feedback outage probabilities for both proposed scheme and the selective scheme are maintained below 3% as targeted in the parameters setting in the previous section. The feedback outage of the proposed scheme is lower than outage of the selective feedback technique which in turn results in a higher sum rate as illustrated in Figure 5.4. The lower feedback outage implies more users would have sent feedback as demonstrated in the simulation results in Figure 5.6.

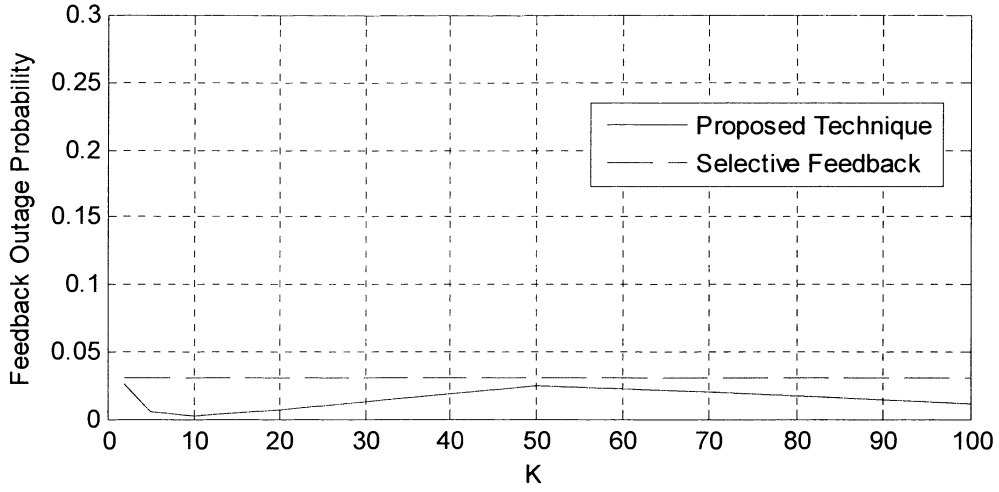


Figure 5.5 Feedback outage probability of the proposed feedback reduction technique

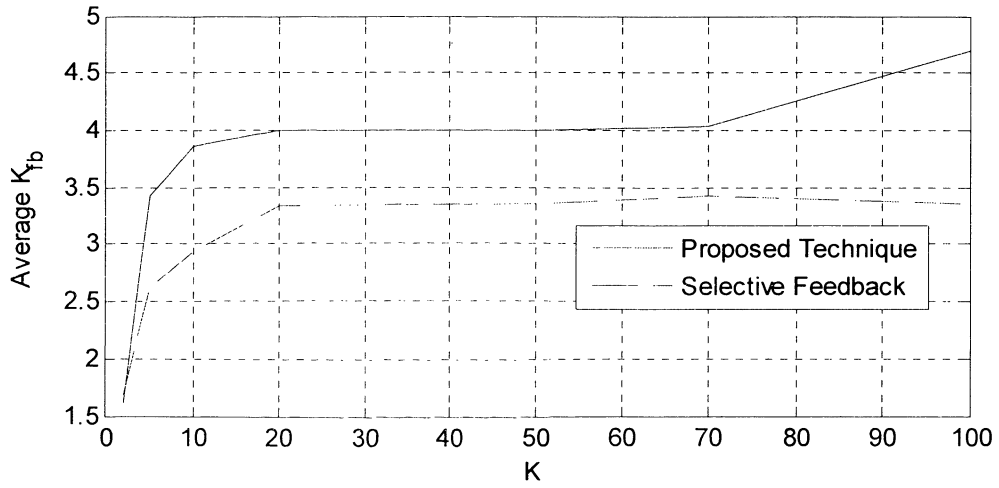


Figure 5.6 Average number of users sending feedback

Figure 5.6 show the average number of users sending feedback for both schemes. It can be observed that both the feedback reduction algorithms can maintain the average number of users sending feedback at below 5 users for K = 2 to 100.

5.2.3 Value of Automatically Adjusted Feedback Threshold

To further illustrate the performance of the proposed scheme, the average values of feedback thresholds are plotted in Figure 5.7. The proposed feedback reduction technique can automatically adjust the feedback threshold similar to Gesbert's work [20] whil the proposed technique does neither require knowledge of the number of users in the system nor does the base station have to continuously calculate and broadcast the feedback threshold.

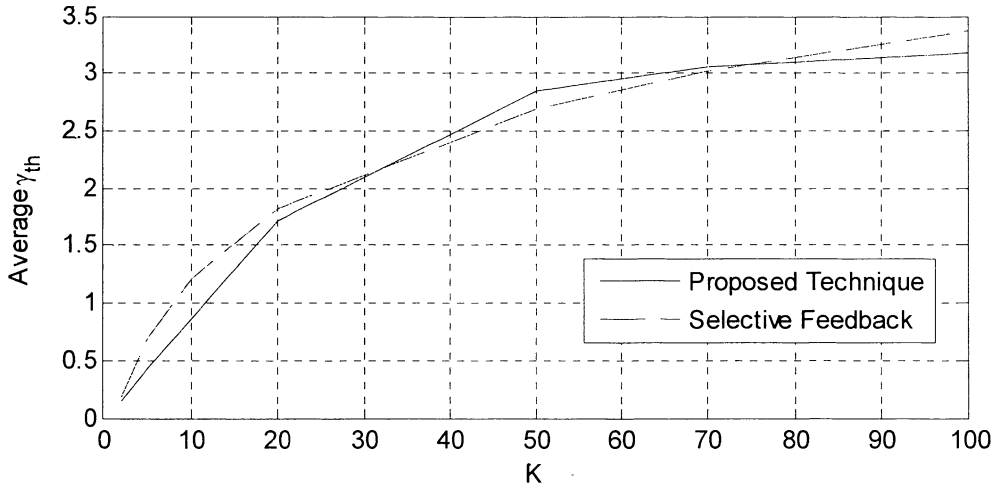


Figure 5.7 The values of feedback thresholds.

5.2.4 Fairness Performance

In this section, the fairness of the proposed scheme is investigated. Figure 5.8 shows the fairness performance indicator of all users for the proposed technique and the selective feedback technique calculated from Equation 5.4. It can be observed that the proposed feedback reduction technique maintain the fairness close to unity.

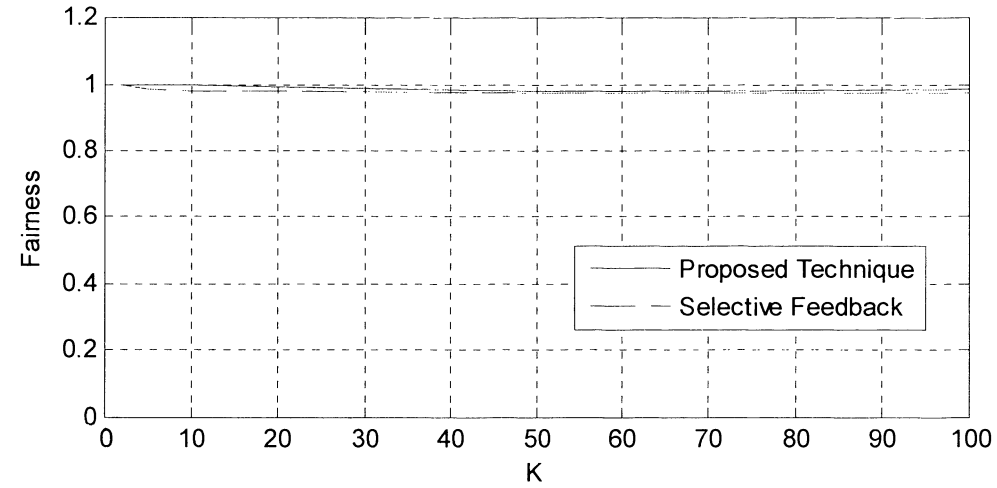


Figure 5.8 Fairness of the proposed feedback reduction technique

5.3 Hybrid Scheduling Policy with Reduced Feedback

In this section, a modified version of opportunistic scheduling algorithm that achieves multiuser diversity gain while minimising the scheduling delay with reduced feedback load is proposed. The proposed technique incorporates the

proportional fairness scheduling policy with the first-in-first-out (FIFO) policy to obtain a hybrid scheduling algorithm that works well when CSI of each mobile user is not necessarily known to the base station.

5.3.1 Scheduling Delay Problem with PF Scheduling

One of the problems inherited from the proportional fair (PF) scheduling policy is that it does not take into account the delay in scheduling. This implies that it may cause excessive delays in a worst-case scenario when the base station has buffers full of packets waiting to be sent to each user. This may result in packets being discarded due to lack of buffering capacity.

Let us consider a multiuser system with K users with PF scheduling. The base station selects a user having the highest relative achievable data rate $[r_k(t)/R_k(t)]$ in each scheduling slot t:

$$k^*(t) = \arg\max_{k=1...K}\left\{\frac{r_k(t)}{R_k(t)}\right\} \tag{5.5}$$

where $r_k(t)$ is the instantaneous achievable data rate and $R_k(t)$ is the long-term average data rate for k^{th} user at scheduling slot t. This scheduling policy can be simplified to:

$$k^*(t) = \arg\max_{k=1...K}\left\{\frac{r_k(t)}{R_k(t)}\right\} \approx \arg\max_{k=1...K}\left\{\frac{\gamma_k(t)}{\bar{\gamma}_k(t)}\right\} \tag{5.6}$$

as previously discussed in Chapter 2.

The cumulative distribution function (CDF) of the scheduling delays, τ, for a system with K =10 and 100 users employing PF scheduling is illustrated in Figure 5.9

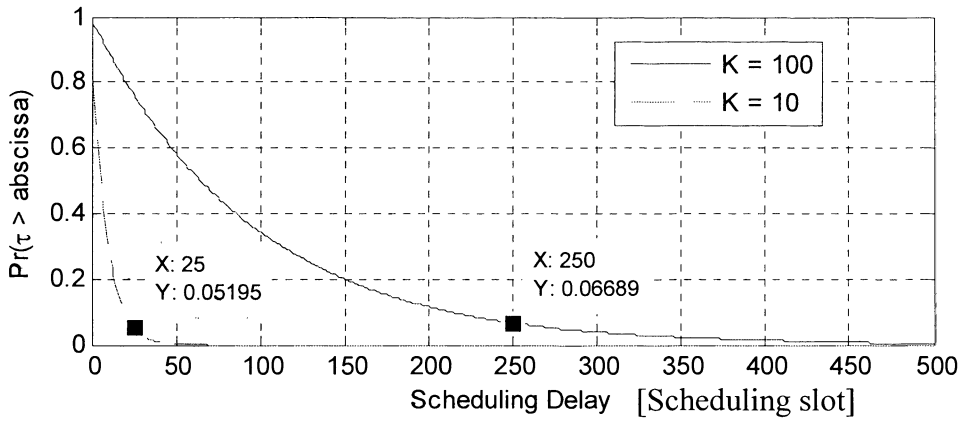


Figure 5.9 CDF of scheduling delays with PF policy

Figure 5.9 shows that the larger the number of users the longer the probability tail of the scheduling delay.

The probability tail shows an important property of packet system with limited queue size. If the scheduling delay exceeds a specific value (i.e. a delay threshold) then the packets in the queues would have expired and been discarded resulting in a performance degradation of the overall system. For the example in Figure 5.9, if the delay threshold is 250ms then 6.68% of data would have been discarded if the PF scheduling was used.

5.3.2 Overview of Delay-Aware Scheduling Algorithms

In this section, scheduling techniques proposed to minimise the scheduling delays will be overviewed.

5.3.2.1 First-in-First-Out (FIFO)

The simplest form of scheduling that can minimise the delays is the first-in-first-out (FIFO) scheduling technique where the radio resource will be allocated to the user with the largest delay. The scheduling policy of FIFO can be expressed as finding the scheduled user, k^*, that:

$$k^*(t) = \arg\max_{k=1...K}(\tau_k(t)) \tag{5.7}$$

where $\tau_k(t)$ is the head-of-line packet delay for user k at the base station at scheduling slot t.

FIFO is a low complexity scheduling technique that does not use knowledge of CSI and thus cannot achieve the multiuser diversity gain.

5.3.2.2 Modified-Largest Weighted Delay First (M-LWDF)

One of the well-known scheduling techniques that effectively minimises the scheduling delays while taking into account the instantaneous channel state information to improve the achievable data rate is a technique called Modified-Largest Weighted Delay First (M-LWDF) [30]. This technique schedules a user based on a scheduling metric $\psi_k(t)$ that is a product of the user's supportable data rate and its delay as calculated by the equation below:

$$k^*(t) = \arg\max_{k=1...K}(\psi_k(t)) \tag{5.8}$$

$$\psi_k(t) = \zeta_k \cdot \tau_k(t) \cdot \frac{r_k(t)}{R_k(t)} \tag{5.9}$$

where $r_k(t)$ is the instantaneous achievable data rate and $R_k(t)$ is the average data rate of user k^{th} at scheduling slot t^{th} respectively. ζ_k is a factor representing the QoS requirement of user k. $k^*(t)$ represents a scheduled user at scheduling slot t.

However, M-LWDF is clearly not optimised to exploit multiuser diversity gain in a system with heavy load where each user's buffer always has packets to be transmitted. More specifically, the scheduling delays, τ_k, for most users except that for the scheduled user in the previous scheduling slot would continuously increase in every scheduling slot (shown later in Figure 5.20). The delay parameter $\tau_k(t)$ in the M-LWDF scheduling policy then dominates the scheduling metric turning the

scheduling policy to behave like FIFO and resulting in diminished contribution of the user's received SNR in the scheduling metric.

5.3.2.3 Exponential Rule (EXPR)

Another well-known scheduling algorithm that was developed for delay-sensitive services is the Exponential Rule (EXPR) algorithm [31]. This scheduling algorithm determines the best user to gain access to the resource based on the following metric $\psi_k(t)$:

$$k^*(t) = \arg\max_{k=1\ldots K}\left(\psi_k(t)\right) \tag{5.10}$$

$$\psi_k(t) = \zeta_k \cdot \frac{r_k(t)}{R_k(t)} \cdot \exp\left(\frac{\zeta_k \cdot \tau_k(t) - \hat{\tau}(t)}{1+\sqrt{\hat{\tau}(t)}}\right) \tag{5.11a}$$

and

$$\hat{\tau}(t) = \frac{1}{K}\sum_{k=1}^{K} \zeta_k \cdot \tau_k(t) \tag{5.11b}$$

$\hat{\tau}(t)$ is defined as the average weighted delay.

This policy aims at scheduling the user based on an exponential term which is a function of scheduling delay. This exponential term would become larger when the difference between the weighted delay (term $\zeta_k \cdot \tau_k(t)$) and the average weighted delay ($\hat{\tau}(t)$) is greater. EXPR technique results in an efficient scheduling for delay sensitive services as demonstrated in the literature. However, in a heavily loaded network, EXPR fails to obtain multiuser diversity gain and thus cannot effectively utilise the knowledge of the received SNR to improve the system performance in terms of achievable data rate as will be demonstrated in Section 5.5. Furthermore, the proposed technique requires users to continuously send CSI to the scheduling thereby consuming feedback resources.

5.3.2.4 Channel-Dependent Earliest Due Deadline (CD-EDD)

In [32], a scheduling algorithm called Channel-Dependent Earliest Due Deadline (CD-EDD) to support delay-sensitive services in cellular networks was described.

This scheduling algorithm allocates a higher priority to a user having delay approaching a packet delay deadline, τ_d. The CD-EDD scheduling policy can be expressed as:

$$k^*(t) = \arg\max_{k=1...K}\left(\psi_k(t)\right) \tag{5.12}$$

and

$$\psi_k(t) = \zeta_k \cdot \frac{r_k(t)}{R_k(t)} \cdot \frac{\tau_k(t)}{\tau_d - \tau_k(t)} \tag{5.13}$$

Although it has been shown that CD-EDD scheduling policy results in excellent performance for delay-sensitive services, it clearly fails to achieve multiuser diversity gain in a heavily loaded network as the $\tau_k/(\tau_d-\tau_k)$ term could dominate the scheduling metric and thus turn this algorithm to a FIFO algorithm.

5.3.3 Hybrid Scheduling with Reduced Feedback

In this section, a new scheduling technique that is a hybrid version of PF scheduling and FIFO (hereafter called HPF: Hybrid FIFO and Proportional fair) is proposed. Similar to the selective scheduling in [20], the HPF algorithm allows the best user to have access to the channel based on PF policy. The difference between the proposed algorithm and the selective feedback algorithm proposed in [20] is the user selection process in feedback outage where none of users sends CSI to the base station. Let us consider the number of feedback users in a given scheduling slot t, $FB(t)$:

$$FB(t) = \sum_k \left(0.5 \cdot \mathrm{sgn}\left(\frac{\gamma_k(t)}{\bar{\gamma}_k(t)} \geq \gamma_{th} \right) + 0.5 \right) \tag{5.14}$$

where sgn is the signum function. The scheduled user, k^* is selected based on the following policies:

$$k^*(t) = \arg\max_{k=1..K}\left(\frac{r_k(t)}{R_k(t)} \right) \quad if\ FB(t) > 0 \tag{5.15}$$

$$k^*(t) = \arg\max_{k=1...K}\left(\tau_k\right) \quad if\ FB(t) = 0 \tag{5.16}$$

where τ_k is the delay of k^{th} user. Equation 5.15 is the typical scheduling policy of PF algorithm while Equation 5.16 is the typical scheduling policy of first-in first-out (FIFO) policy [33]. A flowchart of the proposed algorithm is given in Figure 5.10.

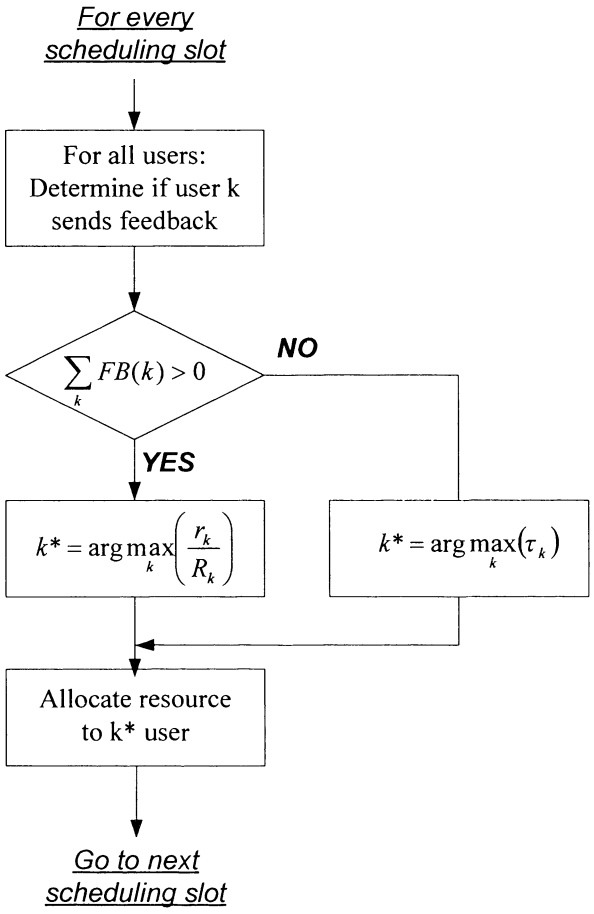


Figure 5.10 Flowchart of the HPF algorithm

The following criterion is used in determining whether a user should send CSI:

$$\frac{\gamma_k(t)}{\bar{\gamma}_k(t)} \geq \gamma_{th} \tag{5.17}$$

where $\frac{\gamma_k(t)}{\bar{\gamma}_k(t)}$ is referred to as relative SNR. The scheduling algorithm can be explained as below:

(i) Each user determines whether to send feedback based on (5.17).

(ii) Base station receives feedback from all users whose relative SNR exceeds the threshold.

(iii) The scheduled user is selected based on PF policy if at least one user sending feedback to the base station; otherwise the user having the large delay is scheduled based on FIFO policy using the delay information available at the base station.

The proposed hybrid algorithm exploits the fact that there is a possibility that feedback outage may occur. In this case, the base station will not have any CSI of any users to properly perform PF scheduling. However, the base station always knows the delay for each user as it is the entity managing data transmission to/from all users. The proposed algorithm allocates the scheduling slot to a user having the highest aggregate delay (head-of-line delay) during a feedback outage. Note that the selective feedback technique in [20] picks a user randomly in the case of feedback outage and that the amount of feedback resource required by the proposed algorithm is the same as that required in [5].

5.3.4 Determining Optimal Feedback Threshold for Hybrid PFS-FIFO

The parameter that affects the performance of the hybrid scheduling algorithm is the feedback threshold which controls the priority if the scheduling would be based mainly on the PF policy or FIFO policy. A number of simulation results are given below to evaluate an optimal value of the feedback threshold. The performance metrics (sum rate and delay probability) are plotted as a function of the feedback outage. The feedback outage can be mapped to the corresponding value of feedback threshold for a given number of users K by the relationship in Equation 5.1 for a given number of users K.

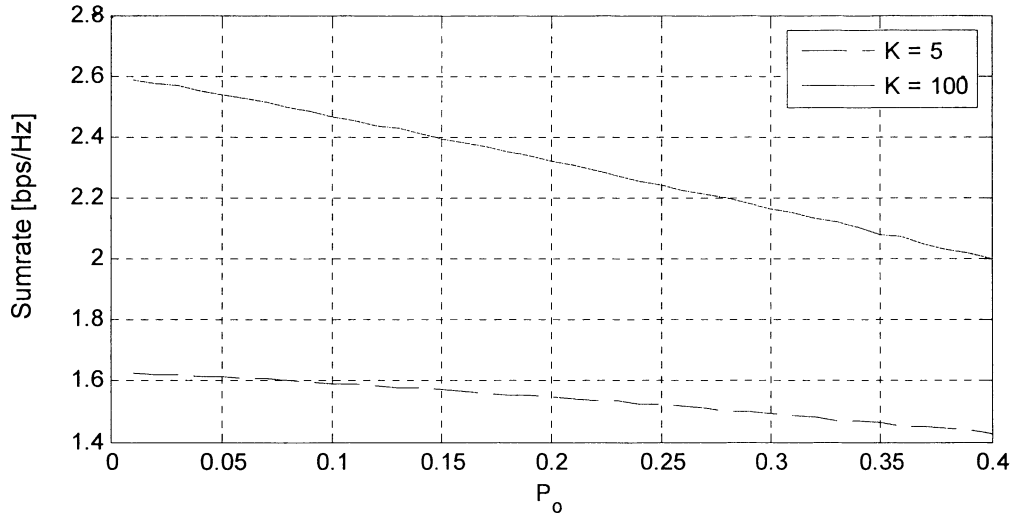


Figure 5.11 Sum rate as a function of feedback outage

Figure 5.11 shows the sum rate performance of the proposed hybrid algorithm as a function of feedback outage for $K = 5$ and $K = 100$. It can be observed that the achievable sum rate drops with an increase of the feedback outage . The optimal value of feedback outage cannot be evaluated from this graph.

Figure 5.12 shows the probability when delays exceed 2.5 times of the average delay as a function of feedback outage for $K = 5$ and $K = 100$.

The maximum acceptable delays is defined as 2.5 times of the average scheduling delays of the system and the probability that the delays exceed this value is P_L. Assume that a packet having a delay greater than this maximum acceptable delay will be dropped. For example, at $P_{ot} = 6\%$, 3.48% of packets would be discarded.

Figure 5.12 shows that with the proposed algorithm, the feedback outage (equivalently feedback threshold) can be adjusted to control the parameter P_L.

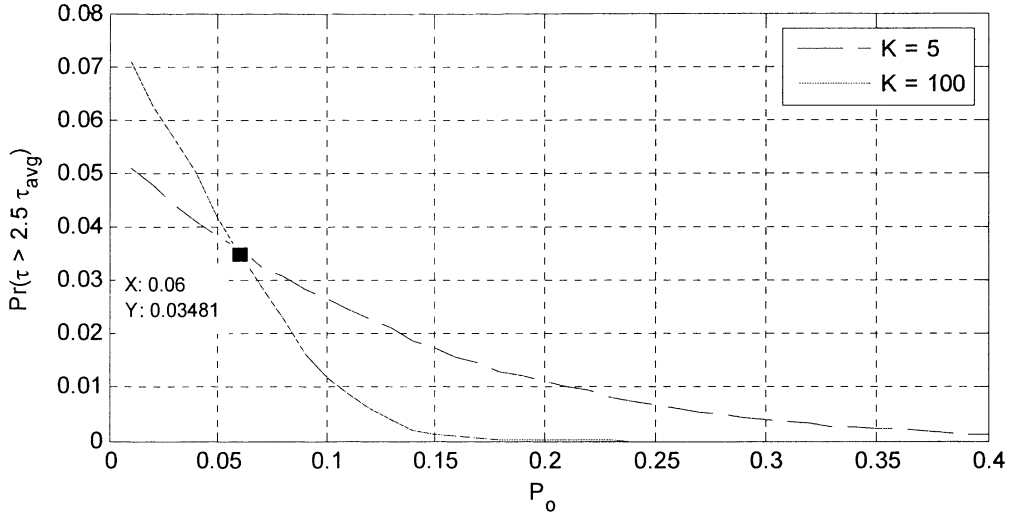


Figure 5.12 Probability that delays exceeds 2.5 times of the average delay

Figure 5.12 demonstrates that the probability P_L decreases exponentially when the feedback outage increases, especially when the number of users is high. At a high feedback outage, the proposed hybrid scheduling algorithm behaves more like FIFO thereby assigning more priority allocating the resource to a user having the largest delays in the queues.

To further optimise the selection of the value of suitable feedback outage, the total achievable rate taking into account the packet loss due to excessive delays should be considered. The total achievable rate, C_{total}, is defined as:

$$C_{total} = (1-P_L) \cdot C_{sum\ rate} \tag{5.18}$$

where $C_{sum\ rate}$ is the achievable sum rate without any loss.

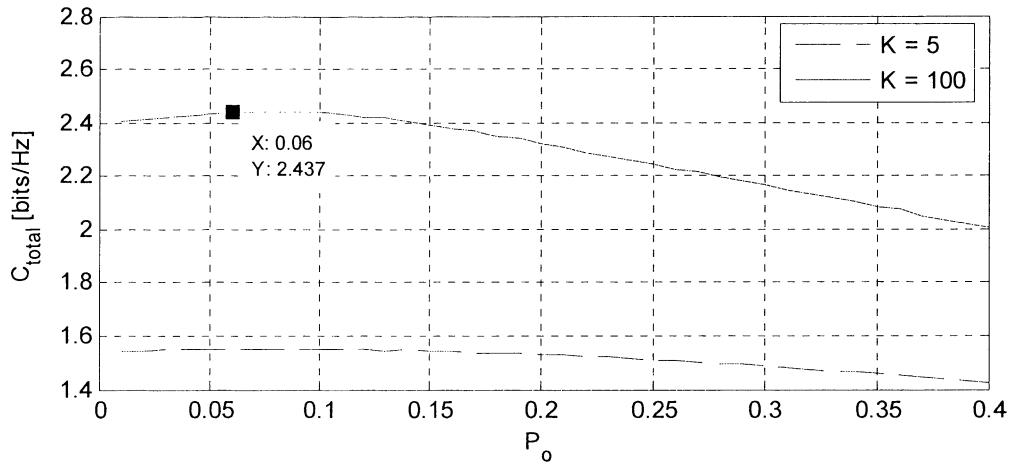


Figure 5.13 Total sum rate taking into account the loss due to excessive delays

Figure 5.13 shows there is an optimal value of the feedback outage that results in a maximum value of the total sum rate capacity ($P_{ot} = 0.06$). This value can be converted to the value of feedback threshold by Equation 5.1 for a given number of users. $P_{ot} = 0.06$ will be used in the subsequent simulations.

5.4 Performance of Hybrid PFS-FIFO

In this section, the performance of the proposed hybrid scheduling algorithm, HPF, is evaluated by simulation. The performance of the proposed HPF is compared with PFS having full CSI feedback.

Figure 5.14 represents the CDF of scheduling delays for the proposed hybrid algorithm and the proportional fair scheduling (PFS) with full CSI feedback.

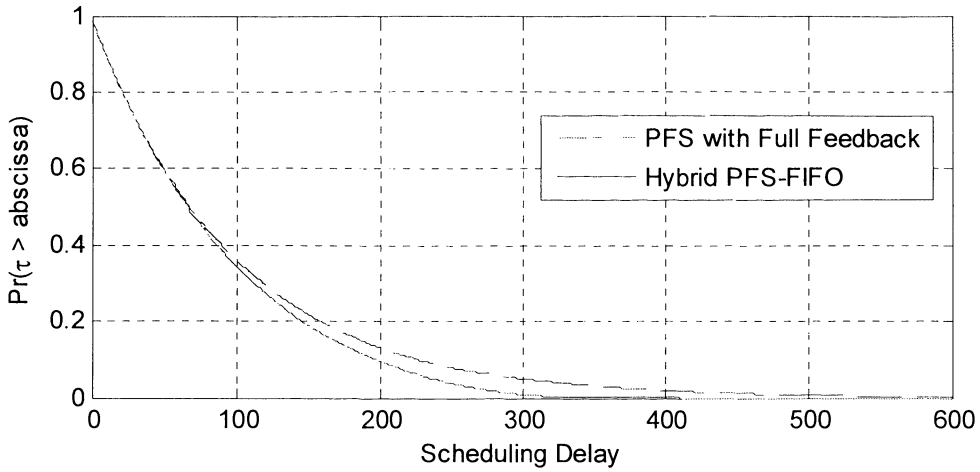


Figure 5.14 Scheduling delays CDF of HPF and PFS

Figure 5.14 shows that the proposed technique can reduce the probability tail of the scheduling delay. This will be beneficial in terms of lower P_L and the lower mean value for the maximum delay as illustrated in the simulation results below.

Figure 5.15 shows the delay performance of the proposed hybrid scheduling with reduced feedback and the PFS with full feedback.

Figure 5.15 indicates that the performance of the proposed technique in terms of the probability P_{ex} that the scheduling delays exceed 2.5 times of the average delay is significantly lower than the P_{ex} of PFS with full feedback.

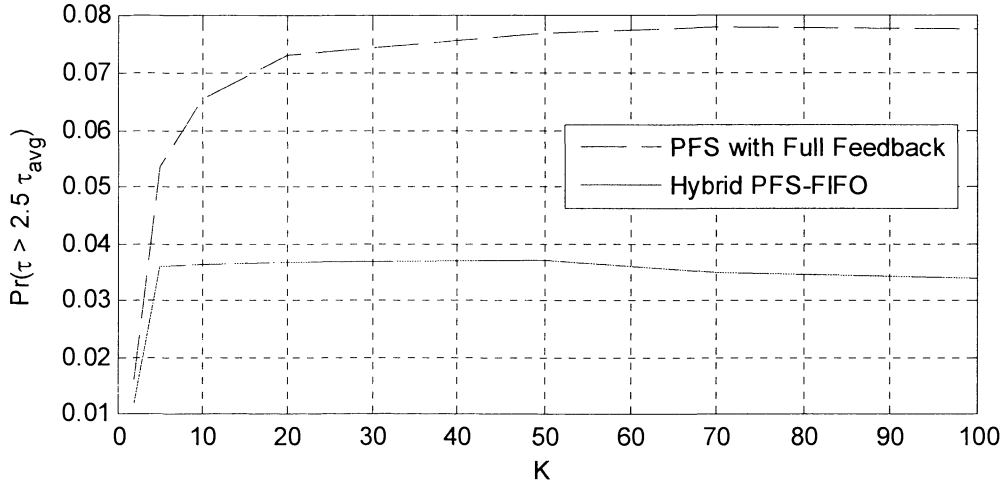


Figure 5.15 Exceed delay probability of HPF and PFS

The proposed technique also reduces the average of the maximum scheduling delay as demonstrated in Figure 5.16.

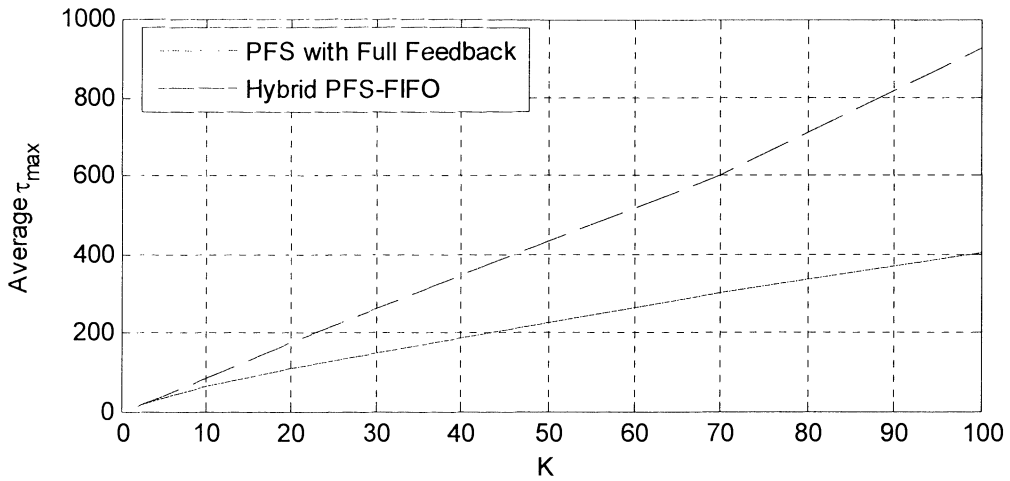


Figure 5.16 Average of maximum scheduling delay for HPF and PFS

It is demonstrated in Figure 5.16 that the mean value of scheduling delay increases linearly with a relatively high rate (approximately 9 scheduling slots per each user counted) when the number of users increases while the mean value of the delays for the proposed hybrid technique increase by a slower rate (i.e. less than 5 scheduling slots per each user counted). The result indicate that the proposed scheduling technique can efficiently reduce the average maximum delay.

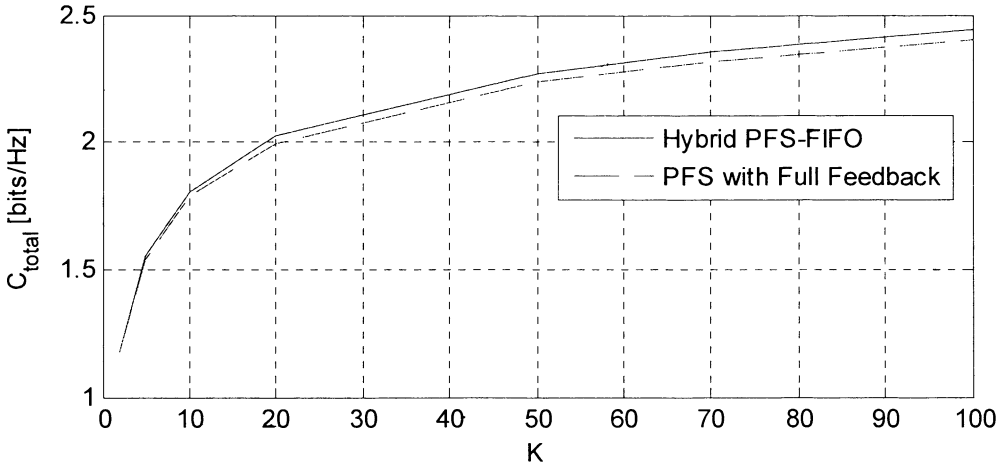


Figure 5.17 Total sum rate of HPF and PFS

Figure 5.17 shows the achievable sum rate capacity of the proposed hybrid feedback algorithm and the PFS. The total sum rate capacity is as defined in Equation 5.18. The HPF can improve the sum rate slightly when compared to PFS with full CSI.

The performance improvement of HPS over PFS is summarised in Table 5.2. The % improvement is defined as:

$$\%\text{Improvement } P_{ex} = 100\cdot(P_{ex,PFS} - P_{ex,HPS})/ P_{ex,PFS} \tag{5.19}$$

$$\%\text{Improvement } \tau_{max} = 100\cdot (\tau_{max,PFS} - \tau_{max,HPS})/ \tau_{max,PFS} \tag{5.20}$$

$$\%\text{Improvement } C_{total} = 100\cdot (C_{total,HPS} - C_{total,PFS})/ C_{total,PFS} \tag{5.21}$$

Table 5.2: Performance improvement of HPS compared with PFS

	K= 2	K= 5	K= 10	K= 20	K= 50	K= 70	K= 100	Average
%Improvement P_{ex}	10.1	6.9	23.4	37	48.8	50.1	56.8	33.3
%Improvement τ_{max}	25	32.4	44.3	49.8	52.1	55.1	56.3	45.0
%Improvement C_{total}	0	0.8	1.2	1.7	1.6	1.8	1.6	1.2

Figure 5.17 together with the results in Figure 5.16 show that the proposed hybrid algorithm with reduced feedback load can achieve higher total sum rate which takes into account the packets that are dropped due to excessive scheduling delay than the PFS with full CSI feedback.

The proposed scheduling technique can significantly reduces P_{ex} and τ_{max} (33.3% and 45% improvement, respectively, compared with PFS as shown in Table 5.2). The results indicate that the performance of PFS technique can be further improved by scheduling users based on scheduling delay parameters in feedback outage.

Figure 5.18 shows the feedback load of the proposed scheme normalised by the load required to send the full feedback. The simulation results in Figure 5.17 and 5.18 demonstrate that the proposed technique can achieve higher total sum rate than PFS with full feedback while requiring substantially less feedback load.

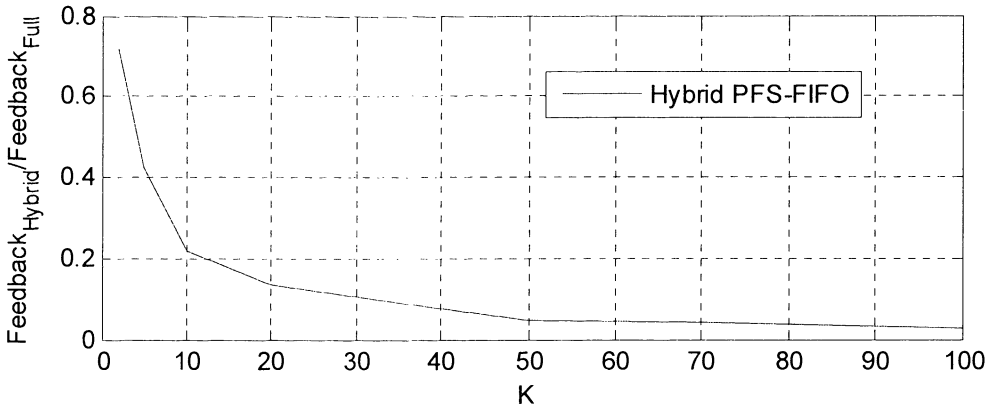


Figure 5.18 Normalised feedback load of the proposed HPF technique

5.5 Performance of Exponential Rule Scheduling in Heavily Loaded Networks

In this section, the performance of the proposed HPF in a heavily loaded network is evaluated and compared with the EXPR. Figure 5.19 shows the total achievable sum rate of the EXPR in comparison with that of HPF and FIFO for K = 2 to 100. The sum rate performance is tabulated in Table 5.3.

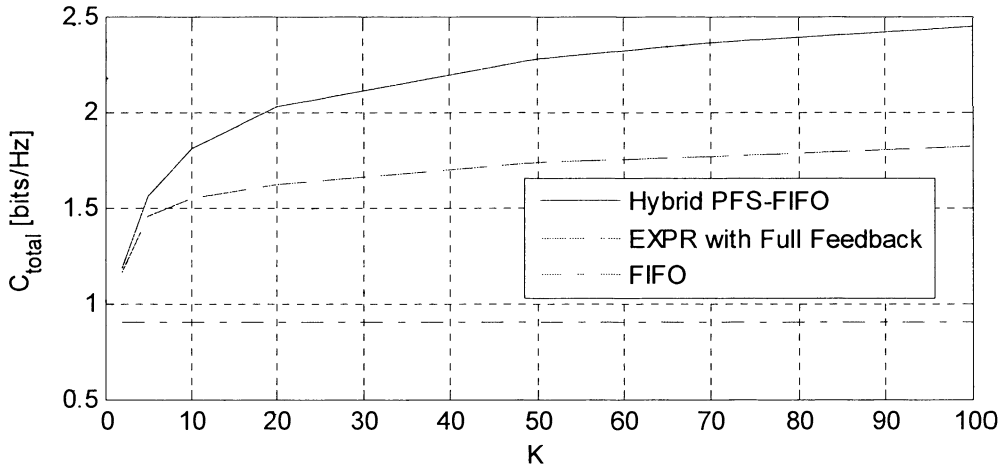


Figure 5.19 Total sum rate of HPF, EXPR and FIFO

Table 5.3 Total sum rate of HPF, EXPR and the improvement of HPF over EXPR.

	K= 2	K= 5	K= 10	K= 20	K= 50	K= 70	K= 100
$C_{total,HPF}$	1.2	1.6	1.8	2.0	2.3	2.4	2.4
$C_{total,EXPR}$	1.2	1.4	1.5	1.6	1.7	1.8	1.8
%Difference	1.3	7.2	16.9	24.8	31.3	33.1	34.3

The simulation in Figure 5.19 demonstrates that HPS significantly outperforms EXPR in heavily loaded scenario. The proposed HPS can achieve more than 30% improvement compared with EXPR when the number of user is higher than 50. This significant difference between HPS and EXPR is due to HPS can still efficiently achieve the multiuser diversity gain while EXPR appears to fail. This is because in the simulation scenario the scheduler's queues are always full of packets to be transmitted. This implies that the scheduling delays of most users would continue to increase in every scheduling slot as illustrated in Figure 5.20 below.

Figure 5.20 represents a snapshot showing the scheduling delays of a network with fully loaded traffic assuming the base station always has packets in the buffer to be transmitted. Different colour for each line represents different users. The scheduling algorithm used in the simulation is PFS. Only 5 users out of 100 are plotted for presentation simplification.

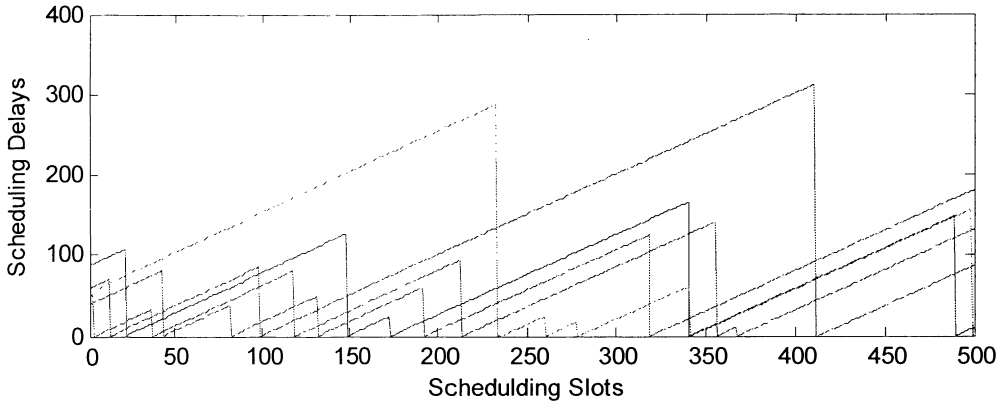


Figure 5.20 Snapshot showing the scheduling delays of a system with 100 users

The scheduling delay in Figure 5.20 is the head of line delay which increases when the packet is placed in the queue and reduces to zero when the packet is scheduled. The delay rise increase linearly one scheduling slot by one scheduling slot.

Figure 5.20 shows that most of the time, the scheduling delays increase continuously in each scheduling slot. This will cause the exponential term in the EXPR to rapidly increase and thus the scheduling allocates more priority to scheduling users with larger delay rather than higher received SNR. As a result, EXPR behaves like FIFO.

Figure 5.21 shows that the CDF's of the scheduling delays for FIFO and EXPR are identical. This validates that EXPR scheduling policy provides more priority in scheduling based on the delays and thus it becomes equivalent to FIFO in a heavily loaded scenario.

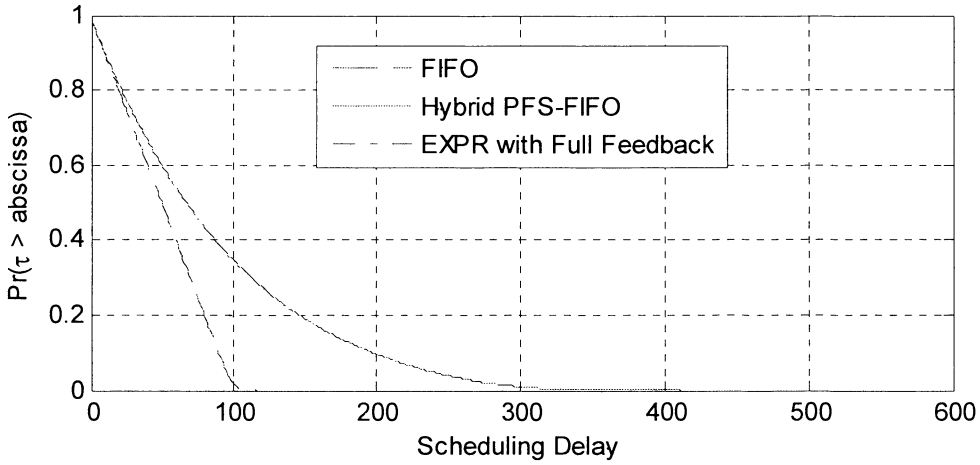


Figure 5.21 CDF of scheduling delays for a system with K =100

The exponential term in Equation 5.10 is plotted in Figure 5.22. It can be observed that the exponential term dramatically increases when the delay approaches the mean value (100 scheduling slots in this case). When the delay exceeds this value, the exponential term of the EXPR increases rapidly and dominates the scheduling metric in Equation 5.10. From the CDF of the scheduling delays in Figure 5.14, there are approximately 30% of users having scheduling delays exceeding the mean delay which implies that these users would have the scheduling metric dominated by the delays and the CSI would have little contribution to the scheduling decision.

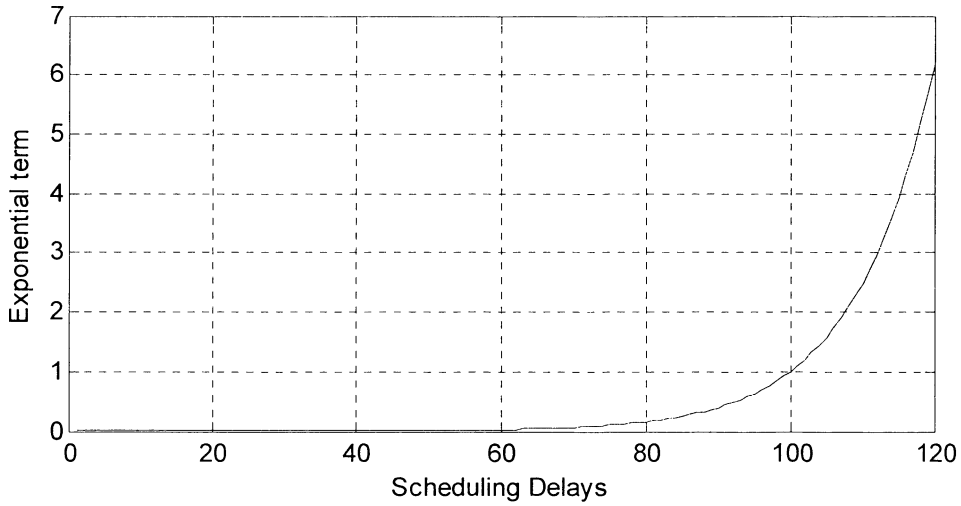


Figure 5.22 Exponential term of EXPR, K = 100

5.6 Implementation Issue

The proposed feedback reduction algorithm with automatic threshold adjustment can be implemented in any existing system implementing the selective feedback algorithm proposed in [20]. The results provide a guideline that the feedback threshold can be automatically adjusted at each user independent of broadcasted message from the base station.

For the proposed hybrid algorithm, provided that the feedback algorithm in [20] has already been used in High Data Rate (HDR) systems, HPF has a complementary functionality to minimise the delay without significantly affecting the achievable sum rate. There is no extra requirement to upgrade the hardware at mobile station to implement the proposed algorithm as only typical CSI is required to be sent. Moreover, there is no requirement to change the broadcast message mechanism to inform all users of the feedback threshold provided that the algorithm in [20] already requires such information. The only modification needed to be made to implement the proposed algorithm in practical is a minor software upgrade at the scheduler i.e. base station.

The proposed hybrid scheduling and the automatic feedback threshold adjustment techniques are generic. They can be applied to most feedback scheduling schemes where the amount of feedback and/or scheduling delays need to be reduced.

5.7 Summary

In the first part of this chapter, a new feedback reduction technique where the feedback threshold is automatically adjusted is proposed. The proposed technique can self-tune the feedback threshold to grow when the number of users increases similar to when calculated from Equation 5.1.

In the second part, a new hybrid scheduling algorithm with reduced feedback, named HPF, which can minimise the scheduling delay is proposed. The proposed algorithm behaves like PFS when one or more users send(s) feedback. If none of the

users send CSI feedback the resources will be allocated to the user having the largest delay (similar to FIFO) by utilising the delay information available at the base station. Simulation results demonstrate that the proposed algorithm can reduce P_{ex} by 33.3%, and τ_{max} by 45% on average compared with PFS. HPF achieves a 1.2% average improvement in the total sum rate taken into account the effect of packets expired due to excessive scheduling delays. The proposed technique outperforms EXPR in the heavily loaded network by more than 30% in term of achievable sum rate. The implementation of HPF does not involve any hardware upgrading.

Chapter 6

Adaptive Feedback for OFDMA Systems

In Chapter 3 and 4, feedback techniques that are designed to reduce the number of bits required to represent all possible MCS modes in single carrier systems were considered. In Chapter 5, an adaptive technique to limit the number of users sending feedback was proposed for single carrier systems.

In this chapter, a new feedback reduction technique for orthogonal frequency division multiple access (OFDMA) systems is proposed. The chapter begins with an introduction to OFDMA system and the achievable sum rate of OFDMA systems is defined in Section 6.1. In Section 6.2, feedback reduction techniques in OFDMA are reviewed. A new feedback reduction technique for OFDMA systems is proposed in Section 6.3. Section 6.4 describes the simulation configuration used in this chapter. Simulation results to demonstrate the performance of the proposed feedback reduction technique are presented and discussed in Section 6.5. Section 6.6 summaries the content of this chapter.

6.1 Introduction

Next generation wireless networks will be designed to support a variety of multimedia applications which require different Quality-of-Service (QoS). OFDMA has been chosen as the radio access technology to support such high requirement applications in modern wireless systems such as WiMAX [34] and Long Term Evolution (LTE) systems [4][10][11]. In OFDMA systems, all users share the radio resource which comprises a number of clusters of resource blocks which exist in time and frequency domains. This enables the use of advanced link adaptation technique while necessitates the need for an efficient resource management and the associated control overhead.

The scheduler (i.e. the base station for the downlink communication) can perform the frequency domain packet scheduling to achieve the multiuser diversity gain. It has been shown that the sum capacity of the system increases as the number of active users becomes large [35]-[37]. However, the allocation in OFDMA systems requires large amount of CSI feedback.

Let us consider an OFDMA system in which the total transmit power at the base station, ρ, is equally distributed over the N subchannels. The achievable data rate of k^{th} user on n^{th} subchannel, $\varepsilon_{n,k}$ can be represented by:

$$\varepsilon_{n,k} = W_n \cdot \mu_{n,k} = W_n \cdot \log_2(1 + \frac{\rho}{N} \cdot \frac{h_{n,k}}{\eta_o}) \tag{6.1}$$

where W_n is the bandwidth of each subchannel (or equivalently each resource block), $\mu_{n,k}$ is the spectral efficiency, $h_{n,k}$ is the subchannel gain, and η_o is the noise level on each subchannel.

The total capacity of the system, C, can then be formulated as the summation of the achievable spectral efficiency of K users over N subchannels i.e.:

$$C = \sum_{k=1}^{K} \sum_{n=1}^{n} I_{k,n} \cdot \mu_{n,k} \tag{6.2}$$

$$C = \sum_{k=1}^{K} \sum_{n=1}^{N} I_{n,k} \cdot \log_2(1 + \frac{\rho}{N} \cdot \frac{h_{n,k}}{\eta_o}) \tag{6.3}$$

where $I_{n,k}$ is the subchannels assignment indicator:

$$I_{n,k} = \begin{cases} 1 & ;if\ subchannel\ n^{th}\ assigned\ to\ k^{th}\ user \\ 0 & ;Otherwise \end{cases} \tag{6.4}$$

subject to:

$$\sum_{k=1}^{K} I_{n,k} = 1 \qquad \forall n,k \tag{6.5}$$

The constraints in Equation 6.4 and 6.5 imply that each subchannel can be allocated to only one user in each scheduling slot.

Equation 6.3 can be rewritten in term of the received SNR at each user on each subchannel, $\gamma_{n,k}$ as below:

$$C = \sum_{k=1}^{K} \sum_{n=1}^{N} I_{n,k} \cdot \log_2(1+\gamma_{n,k}) \tag{6.6}$$

where

$$\gamma_{n,k} = \frac{\rho}{N} \cdot \frac{h_{n,k}}{\eta_o} \tag{6.7}$$

It has been shown in the literature [35]-[37] that allocating each subchannel to a user who has the highest received SNR in each scheduling slot maximises the achievable sum rate. This necessitates accurate channel state information (CSI) to be available at the base station so that the subchannel selection and adaptive modulation and coding scheme (MCS) can be performed effectively [38]-[41].

However, the amount of the feedback required to perform a subchannel allocation and adaptive MCS could be prohibitive to the overall system capacity because the amount of feedback resource grows linearly as a function of the total number of users sending feedback K_{fb}, number of subchannels N, and MCS levels L, as expressed below:

$$FB_{full} = K_{fb} \cdot N \cdot log_2(L) = K_{fb} \cdot N \cdot B \tag{6.8}$$

where B is the number of bits to represent MCS level i.e. $B = log_2(L)$.

6.2 Overview of Feedback Reduction for OFDMA Systems

Several feedback techniques have been proposed in the literature to minimise the amount of feedback resources required to achieve multiuser diversity gain in OFDMA systems.

6.2.1 Best M Feedback

In [43], the authors introduced a simple feedback reduction technique for OFDMA system in which the feedback message contains only the CSI from the best M subchannels out of the total N subchannels. The proposed best-M technique provides an acceptable performance in terms of achievable throughput while it offers a significant reduction in the feedback overhead. This technique has already been adopted in WiMAX [44] and been included into LTE specifications [11][44][45]

The total feedback load of the best-M scheme can be formulated as:

$$FB_{best\text{-}M} = K_{fb} \cdot M \cdot (\log_2(L \cdot N)) = K_{fb} \cdot M \cdot (\log_2(N) + B) \qquad (6.9)$$

The performance of the best M system with 25 subchannels is studied in Figure 6.1.

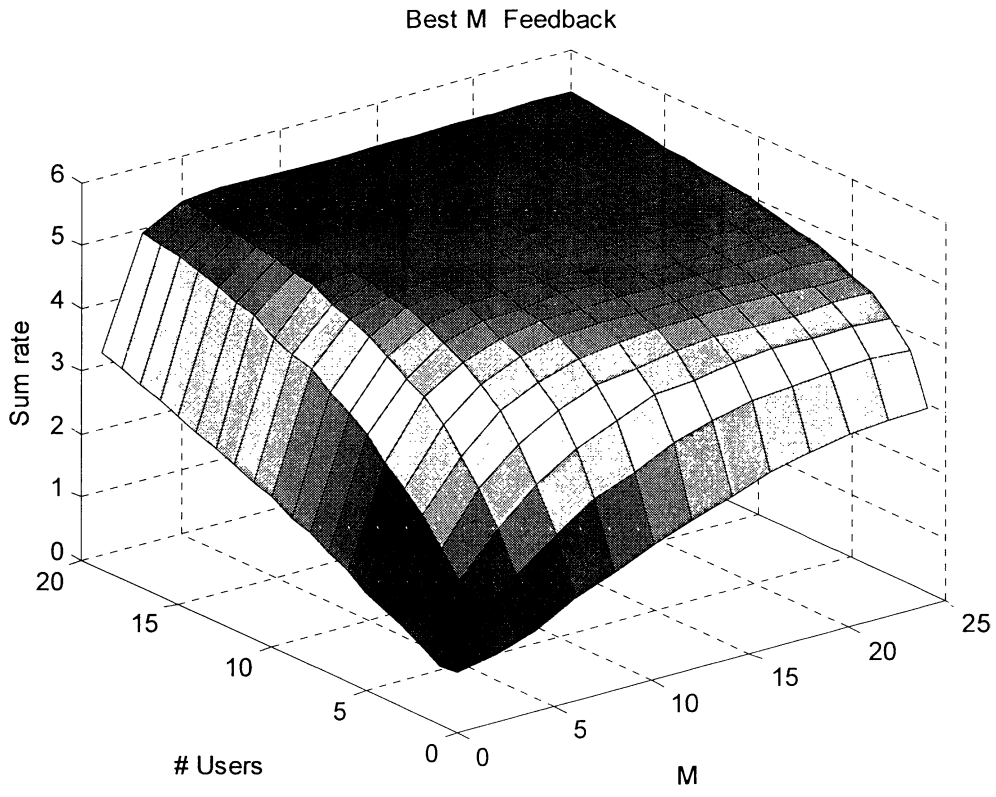


Figure 6.1 Achievable sum rate of multiuser system with best M feedback

Figure 6.1 shows the achievable sum rate (in bps/Hz) of a multiuser system with 25 subchannels versus the number of users K and number of subchannels to report CSI, M. Refer Section 6.3 for a description of simulation model.

Figure 6.1 indicates that the performance of best-M technique varies depending on the values of M and the number of users K. More specifically, the best-M technique causes relatively large performance drop when the number of users is small. This is due to the feedback outage where none of users reports CSI for a subchannel. The feedback outage for best-M technique is expressed as [46]:

$$P_{ot,bestM} = \left(1 - \frac{M}{N}\right)^K \tag{6.10}$$

and the outrage probability in Equation 6.10 is plotted in Figure 6.2 below:

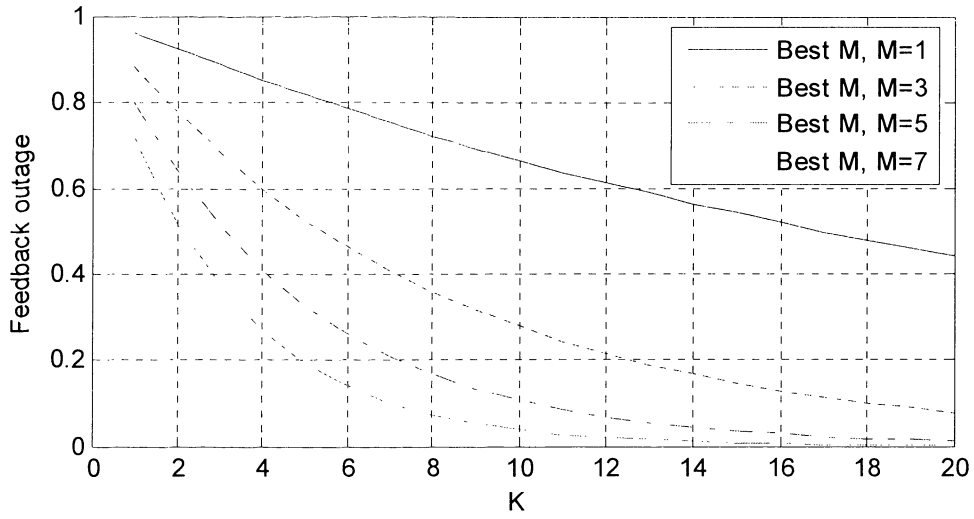


Figure 6.2 Outage probability of best M feedback

Figure 6.2 shows that there is a high probability of feedback outage when the number of users is low. For example, at K =2 and M = 5, the feedback outage probability exceeds 60% which indicates that the system is likely to have no CSI information to perform an efficient scheduling. The outage probability decreases rapidly when the number of users increases. This implies that the performance of the best M technique would be close to full feedback when the number of users is high which can be seen from Figure 6.1 at a high value of M.

In [47], it was suggested that the amount of bits required for best M feedback technique can be further reduced to:

$$FB_{best\text{-}M'} = K_{fb} \cdot (M \cdot B + log_2 \binom{N}{M}) \qquad (6.11)$$

where $\binom{N}{M}$ is the number of different combinations of M CSI reporting subchannels from the total of N subchannels.

6.2.2 One-Bit Feedback Reduction

In [48][49], the concept of feedback threshold as discussed in [21][22] was extended from single-carrier to OFDMA systems. With the feedback reduction technique, each user determines the received SNR of each subchannel whether it is above the feedback threshold. The CSI message is constructed such as bits '0' represent subchannels having SNR lower than the threshold and bits '1' otherwise. This technique requires the base station to allocate optimally selected feedback threshold to suit system conditions (e.g. number of users). The amount of feedback required for the one-bit feedback for OFDMA can be represented by:

$$FB_{one\text{-}bit'} = K_{fb} \cdot (N + B) \qquad (6.12)$$

6.2.3 Average Best M Feedback

The amount of feedback resource required can be further reduced by the use of a feedback reduction technique called "average best M" [50]. This technique creates a CSI message containing a subchannel bitmap that represents best M subchannels. The average SNR value of all M subchannels is represented by a single quantised value. Hence, the number of feedback bits required can be expressed by:

$$FB_{A\text{-}best\text{-}M} = K_{fb} \cdot (log_2 \binom{N}{M} + B) \qquad (6.13)$$

which has significantly reduced from the previously scheme.

The performance of average best M feedback reduction technique relies on the selections of the value of M. Figure 6.3 shows the achievable sum rate of average best M for various M and K. The details of the simulation model are as given in Section 6.3.

Figure 6.3 indicates that average best M is highly sensitive to the choice of M. More specifically, for low K, the highest sum rate occurs approximately at M = 12 for K = 2 while the peak sum rate for high number of users appears at much lower M i.e. at M = 3 for K = 20.

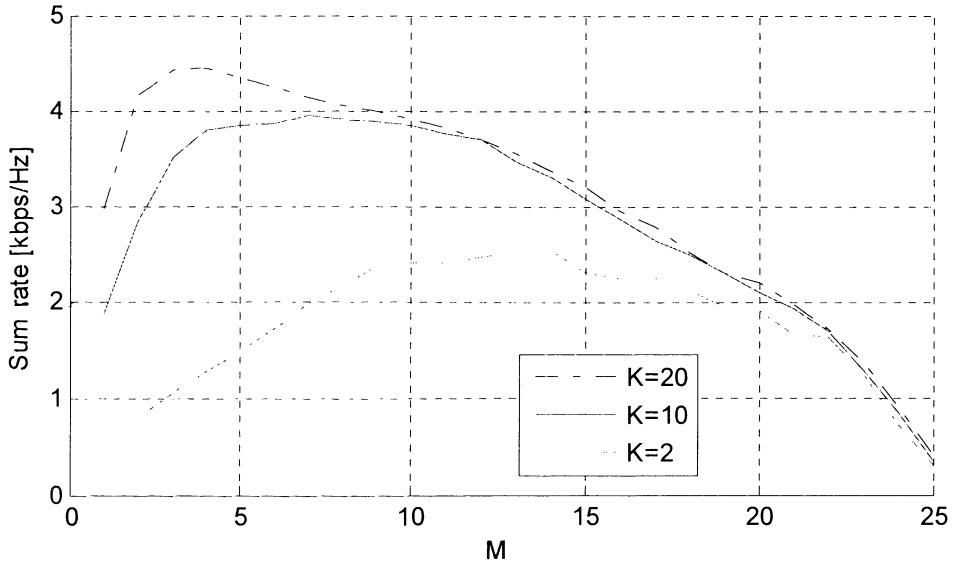


Figure 6.3 Achievable sum rate of average best M feedback reduction technique

Figure 6.3 clearly demonstrates that there is an optimal M which depends on the number of users, K. For K = 2, the optimal value is approximately at M = 12. This value is optimal because at this point the feedback CSI outage is minimised (see Figure 6.2) and at the same time provides an acceptable indication of the subchannels' SNR. At a higher M, the feedback message would no longer capture an accurate estimation of subchannels' SNR due to the averaging process. In contrast, if M was set too low, there will be a large number of subchannels where SNR was not reported. For example, for K = 2, and M = 5, the maximum number of subchannels in which CSI may be reported is $2 \cdot 5 = 10$ subchannels out of 25 total

subchannels. The significant loss is due to the missing CSI on the remaining 15 subchannels.

For a relatively large number of K i.e. K = 20, the optimal M is at a lower value, approximately M = 3 in Figure 6.3. This is because there would be many users already reporting CSI (3·20 = 60 which means roughly 2.4 users would have been reporting SNR on each subchannel). Therefore, a smaller M which could capture more accurate average SNR among M reporting subchannels would be more beneficial to the system. On the other hand, if M is set at an inappropriate large value, the performance significantly drops. This is because the CSI would be an inaccurate indication of the SNR of M subchannels and could not represent the peak of the received SNR. As a result, the gain from frequency diversity and multiuser diversity diminishes.

The average best M creates a large performance gap at a relatively low K when the value of M is small and at high K when M is large. Although it is possible to simply adjust the value of M based on the number of users, this would require extra overhead between the users and base station. This would introduce additional complexity to the base station to determine the optimal M in realtime and broadcast to all users.

The best M feedback technique as discussed in Subsection 6.3.1 especially when the number of users is low imposes challenges to develop a feedback reduction technique that works well at the relatively low number of users while exploits the multiuser diversity gain when the number of users increase. The focus of this chapter is to develop such a feedback reduction technique.

6.3 Proposed Adaptive Feedback Reduction Technique

In this section, a feedback scheme is proposed for downlink OFDMA systems with frequency domain packet scheduling to achieve multiuser diversity. The proposed scheme automatically increases M when the number of users is low as it will minimise the probability of feedback outage. When the number of users increases,

the proposed scheme automatically reduces the value of M which will increase the probability of selecting the best subchannel(s).

The characteristics of the new feedback reduction technique are summarised as follows:

1. With the proposed feedback scheme, the value of the feedback bit-mask is automatically adjusted to suit the number of users without the need for the base station to select an optimal value of M to suit the number of users K and thus there is no extra overhead between users and base station.

2. Unlike the one-bit feedback [48][49], there is no need to optimise the feedback thresholds, which is computationally expensive.

3. The acknowledgement messages can be used to estimate the mean SNR value for each user eliminating the need to transmit B bits to represent the average quality of the M reporting subchannels.

The proposed scheme contains two parts: CSI reporting and modified adaptive MCS. The detail of each function of the proposed adaptive feedback scheme is described in the following subsections:

6.3.1 CSI Reporting Scheme

The procedure of the proposed feedback scheme is presented in Figure 6.4.

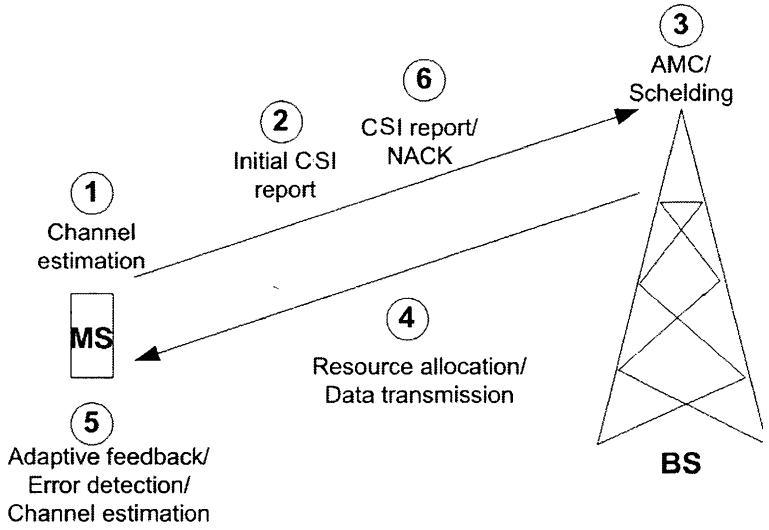


Figure 6.4 The procedure of the proposed feedback scheme

Step 1: Each user estimates the received SNR (assume perfect estimation).

Step 2: Initially, each user sends a CSI message to the base station containing a bit-mask indicating best M_k subchannels (N bits) along with the quantised SNR message (B bits). This M_k value will be dynamically adjusted by the proposed adaptive feedback scheme which will be discussed in Step 5 below.

Step 3: The base station allocates the appropriate radio resources and determines an optimal *MCS* for each subchannel based on CSI received from the users. The method to determine an optimal *MCS* was described in Section 6.3.2 below.

Step 4: The base station transmits the data based on the resources allocated to each scheduled user.

Step 5*:* Each user dynamically determines the CSI feedback message by adjusting M_k based on the following equation (the index k is omitted for notation simplicity):

$$\begin{aligned} M &= \lfloor M + \Delta M_u \rfloor ; \quad if\ M_A = M \\ M &= \lfloor M - \Delta M_d \rfloor ; \quad if\ M_A < M \end{aligned} \tag{6.14}$$

where M_A is the number of allocated subchannels, ΔM_u and ΔM_d are used to increase or descrease M . $\lfloor x \rfloor$ is the highest lower integer of x. The proposed feedback scheme will increase the value of M for a user when the base station is able to schedule on all subchannels indexed on the CSI bit-mask received from that user. The optimal selection of ΔM_u and ΔM_d will be discussed in Section 6.4.

Step 6: Each user reports a CSI feedback message to the base station. Note that the CSI message contains only N bits, not $N+Q$ bits as in Step 2.

Thereafter Step 3 – 5 are repeated.

6.3.2 MCS with Reconstruction of SNR from ACK/NACK

In this section, the second part of the proposed technique is to determine the supportable MCS for each user from ACK/NACK messages.

Adaptive modulation and coding scheme (AMC) is responsible for selecting an optimum MCS to suit the current channel conditions and BLER constraint. At the beginning of each data transmission session (Step 2), each user is required to send a discrete SNR, Γ, so that the base station can determine the optimal MCS. After this beginning stage, the base station determines the optimal MCS based on information extracted from ACK/NACK message. In particular, the base station can construct the Γ by using ACK/NACK messages as following:

$$\delta_k(t) = \delta_k(t-1) + \Gamma_{offset\,k} \quad ; when\, t > 1 \tag{6.15}$$

and

$$\delta_k(1) = \frac{1}{M} \cdot \sum_{m=1}^{M} \Gamma_{k,m} \quad ; when\, t = 1 \tag{6.16}$$

where $\delta_k(t)$ is the quantised SNR extracted from ACK/NACK message, and t is the scheduling slot. $\Gamma_{k,m}$ is the discrete received SNR of m^{th} subchannel of k^{th} user. The parameter $\Gamma_{offset\,k}$ is an offset value of k^{th} user to determine an average value of received SNR. Detail of this offset parameter will be discussed in the next section.

The proposed algorithm exploits HARQ messages to aid link adaptation technique (LAT). The situation can be depicted in Figure 6.5.

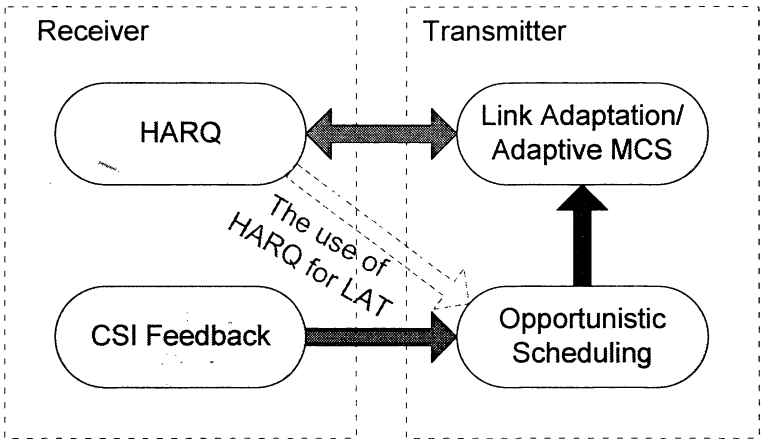


Figure 6.5 Use of HARQ for LAT

For the proposed adaptive feedback technique, in contrast to the scheme in [50], there is no need to send a message representing the quantised SNR (B bits) to represent the average SNR of the M best subchannels in every scheduling slot. In this case, the CSI feedback message consists of N bits which contains a bit-mask indicating the locations of M best subchannels only.

6.4 Simulation Descriptions

The downlink data transmission of an OFDMA system having N subchannels is considered. The total transmit power is constant at P_t. A frequency-selective Rayleigh fading channel is considered in this work. It is assumed that the channel gains for all users are independent and identically distributed (i.i.d.) for all subchannels. For user k on subchannel n, the channel gain $h_{k,n}$ is assumed to be perfectly known at the respective user. The number of subchannels is $N = 25.$. The received signal-to-noise ratio (SNR) of user k on subchannel n is defined as $\gamma_{k,n} = |h_{kn}|^2/\sigma^2$ where σ^2 denotes the variance of the zero-mean additive white Gaussian noise (assumed $\sigma^2 = 1$). The received SNR of the users are uniformly distributed from $10\,dB$ to $20\,dB$.

Adaptive modulation and coding (AMC) technique is implemented to determined an optimal modulation and coding scheme (MCS) mode for the quantised SNR, $\Gamma_{k,n}$ subchannel.

The data rate, $r_{k,n}$, and the block error rate (BLER) is determined by [51].

$$r_{k,n} = mcs \text{ when } \Gamma_{k,n} \in [\gamma_{mcs}, \gamma_{mcs+1}) \tag{6.17}$$

and

$$BLER_{mcs}(\Gamma) = \begin{cases} 1 & ; \ if \ 0 < \Gamma < \gamma_{pn} \\ a_{mcs} \exp(-g_{mcs} \cdot \Gamma) & ; \ if \quad \Gamma \geq \gamma_{pn} \end{cases} \tag{6.18}$$

where values of a_{mcs}, g_{mcs}, and γ_{pmcs} are as defined in Table 6.1 [51].The achievable sum-rate of the system can be expressed as:

Table 6.1 MCS modes [51]

	Mode 1	Mode 2	Mode 3	Mode 4	Mode 5	Mode 6	Mode 7
Modulation	BPSK	QPSK	8-QAM	16-QAM	32-QAM	64-QAM	128-QAM
Rate(bits/sym.)	1	2	3	4	5	6	7
a_n	67.7328	73.8279	58.7332	55.9137	50.0552	42.5594	40.2559
g_n	0.9819	0.4945	0.1641	0.0989	0.0381	0.0235	0.0094
γ_{pn}(dB)	6.3281	9.3945	13.9470	16.0938	20.1103	22.0340	25.9677

$$C_k = \sum_{k=1}^{K} \sum_{n=1}^{N} I_{k,n} \cdot r_{k,n} \tag{6.19}$$

where $I_{k,n}$ is the subchannel assignment indicator:

$$I_{k,n} = \begin{cases} 1 & ; if \ subchannel \ n^{th} \ assigned \ to \ k^{th} \ user \\ 0 & ; Otherwise \end{cases} \tag{6.20}$$

The base station assigns the user to be scheduled on each scheduling time slot based on the modified proportional fair policy. When two or more users appear to have the same scheduling metric on the same subchannel, the base station selects a scheduled user in random. A feedback outage occurs when all users on a subchannel report bit-mask index values '0'. In this case, the base station has no knowledge about the channel condition on this subchannel, and a user will be selected in random.

The BLER target is set at 0.01. When an erroneous transmission occurs, the received data is discarded and a negative-acknowledgement (NACK) message is sent to the base station. Otherwise, an acknowledgement (ACK) message will be sent.

The base station may know an approximate SNR using the CSI sent from the users only while the user knows the accurate received SNR, $\gamma_{k,n}$,. The word size of an initial CSI report, CSI_{init}, which will be sent at the beginning of each session, is given by :

$$CSI_{init} = N + B \text{ bits.} \tag{6.21}$$

where B is the number of bits to represent the quantised SNR levels. N is the number of bits to represent the subchannel bit-mask message($b_{k,n}$) and e $\gamma_{k,n}$ is within the M best quality subchannels:

$$b_{k,n} = \begin{cases} 1 & ; if\ \gamma_{k,n}\ is\ within\ M\ best \\ 0 & ; Otherwise \end{cases} \tag{6.22}$$

The approximate SNR from the channel estimation is converted to discrete SNR values by [47]:

$$\Gamma_{dB} = Q_{StepdB} * floor\left(\frac{\gamma_{dB}}{Q_{StepdB}} + 0.5 + \Gamma_{offset} \right) \tag{6.23}$$

where Q_{StepdB} is the quantisation step in dB (Q_{StepdB} =1dB in this chapter). Γ_{offset} is the offset SNR which is determined by the outer loop link adaptation (OLLA, [52]). The value of the offset SNR is determined as follows:

$$\Gamma_{offset} = \begin{cases} \Delta\Gamma_{up} & ; if \quad ACK\ is\ detected \\ -\Delta\Gamma_{down} & ; if\ NACK\ is\ detected \end{cases} \tag{6.24}$$

In this chapter, the value of B is set at 5 bits, which represents a SNR reporting range of 32 dB (B = 32) with 1-dB quantisation step. According to the average best M feedback, the base station has the knowledge of M subchannels and that the

average SNR of these subchannels is Γ. $\Delta\Gamma_{up}$ and $\Delta\Gamma_{down}$ is set at 0.01 and 0.99 respectively to meet BLER target as per method mentioned in Ref [50].

For the average best M scheme, the optimal M is obtained from finding M that satisfies the following formula for each number of users:

$$M_{op}(K) = \arg\max_{M}(C_K(M)) \tag{6.25}$$

where $C_K(M)$ is the sum rate of average best M.

Similarly, the optimal γ_{th} is obtained from searching solutions that satisfy:

$$\gamma_{th,op}(K) = \arg\max_{\gamma_{th}}(C_K(\gamma_{th})) \tag{6.26}$$

where $C_K(\gamma_{th})$ is the sum rate of one-bit technique.

6.5 Simulation Results

The performance of the proposed adaptive feedback scheme is evaluated under Monte Carlo Simulation using MATLAB. The sum-rate capacity of the proposed scheme is studied in comparison with other limited feedback algorithms namely average best-M [50] and one-bit per subchannel [48][49] feedback algorithms. Simulation parameters are summarised in Table 6.2.

Table 6.2 Simulation parameters for the proposed adaptive feedback technique

Descriptions	Proposed Adaptive feedback	Average Best M	One-bit feedback
Average SNR	Uniformly distributed $\gamma_k \in [10,20]$dB		
Users requirements	Homogeneous (same QoS requirement)		
Number of subchannels	25		
Scheduling policy	Simplified PF		
Scheduling slot	1ms		
Channel model	Uncorrelated Rayleigh		

BLER target	0.01 [8]		
SNR report	B = 5 (32 SNR quantisation levels)		
MCS Modes	7 modes (refer Table 6.1)		
Update parameters when K changes	Self adjustment	Update M by BS	Update γ_{th} by BS
MCS selection	By HARQ messages (no CSI required)	By CSI	By CSI
CSI bits required	N	$N + B$	$N + B$
ΔM_u	1 dB	N/A	N/A
ΔM_d	0.01 dB	N/A	N/A

Figure 6.6 shows the sum-rate of four scenarios: the proposed adaptive feedback, average best-M with M = 5 and M =10, and best M (M=3). All performances curves have been normalised by the sum-rate of the system with full CSI reporting, C_{full}. The normalised sum-rate is defined as C/C_{full}. Best M with M = 3 is used as it requires approximately the same number of feedback bits (27 bits) to the proposed feedback scheme (25 bits).

The achievable sum rate normalised by the full-CSI sum rate and the sum rate improvement (in %) when compared with average best M with M = 5 and best M with M = 3 is given in Table 6.3.

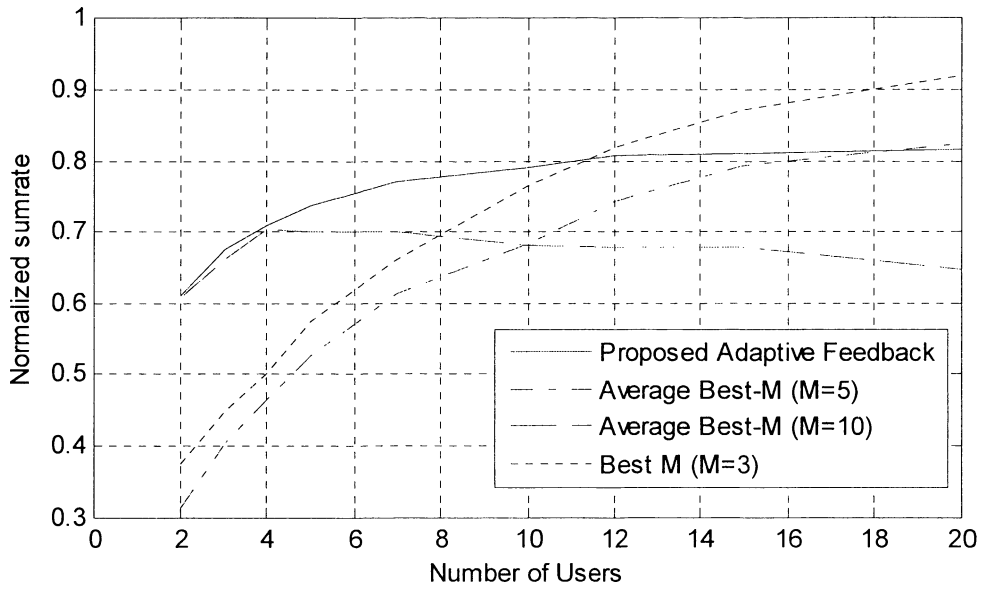


Figure 6.6 Performance comparison with average best M (fixed M)

[8] A small value of BLER is selected as it would limit the retransmission and would be suitable for real time traffic

Table 6.3 Summary of performance improvements of the proposed scheme vs average best M (M = 5) and Best M (M = 3)

Number of user, K	2	3	4	5	7	10	12	15	20	Average
Proposed adaptive feedback compared with full CSI	61.1%	67.5%	70.9%	73.6%	77.1%	79.1%	80.7%	80.9%	81.7%	74.7%
% Improvement compared with Average Best M (M=5)	29.7%	27.2%	24.3%	20.8%	15.8%	10.6%	6.5%	1.6%	-0.8%	15%
% Improvement compared with Best M (M=3)	23.6%	22.8%	20.8%	16.1%	11.1%	2.45%	-1.3%	-6.3%	-10.3%	8.8%

From Figure 6.6, it can be observed that the performance of the proposed adaptive feedback scheme outperforms average best M with fixed M for M = 5 and 10. Furthermore, the proposed scheme outperforms the best M with M = 3 when the number of users is smaller than 11. This is because, at a high value of K, best M scheme can determine more accurate CSI than the average of M best subchannels due to the averaging process of the average best M.

From Table 6.3, the proposed adaptive feedback technique can provide a sum rate improvement of approximately 15% for K = 2 to 20 when compared with the average best M. A sum rate improvement of around 8% can be obtained when compared the sum rate of the proposed feedback reduction technique with best M (M = 3) technique. It can be observed that the proposed technique significantly outperform average best M and best M techniques when the number of users is small (20.8% and 16.1% improvement respectively, when K = 5).

The performance of the proposed feedback scheme is compared with the performance of average best M with optimally selected M and one-bit feedback with optimally adjusted feedback threshold in Figure 6.7.

In Figure 6.7, the normalised sum-rate of the proposed scheme matches the normalised sum-rate of the average best M feedback with optimally selected M.

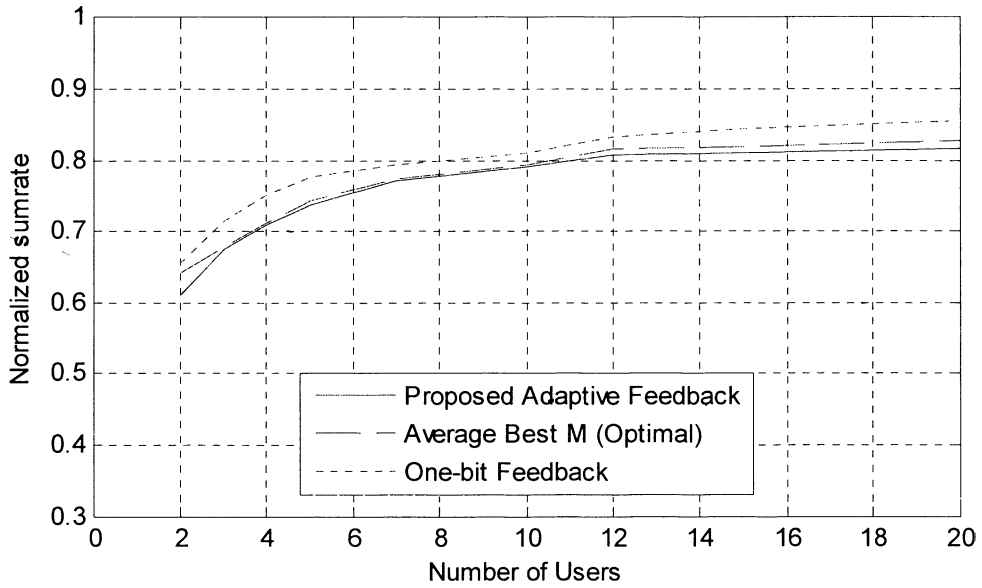


Figure 6.7 Comparison of normalised sum-rate

The result in Figure 6.7 implies that the proposed feedback scheme is capable of automatically adjusting the parameter M to suit the number of users without any requirement for extensive searching for an optimal M.

The performance of the proposed feedback scheme is slightly lower than the normalised sum-rate of one-bit feedback algorithm with optimally selected feedback thresholds. This is because the bit-mask of one-bit feedback is determined by optimally selected feedback threshold and it could more accurately represent the SNR of the best subchannels than averaging from M reporting subchannels.

Figure 6.8 shows the comparison between the average value of the parameter M of the proposed adaptive feedback and the optimally selected parameter M of the average best M.

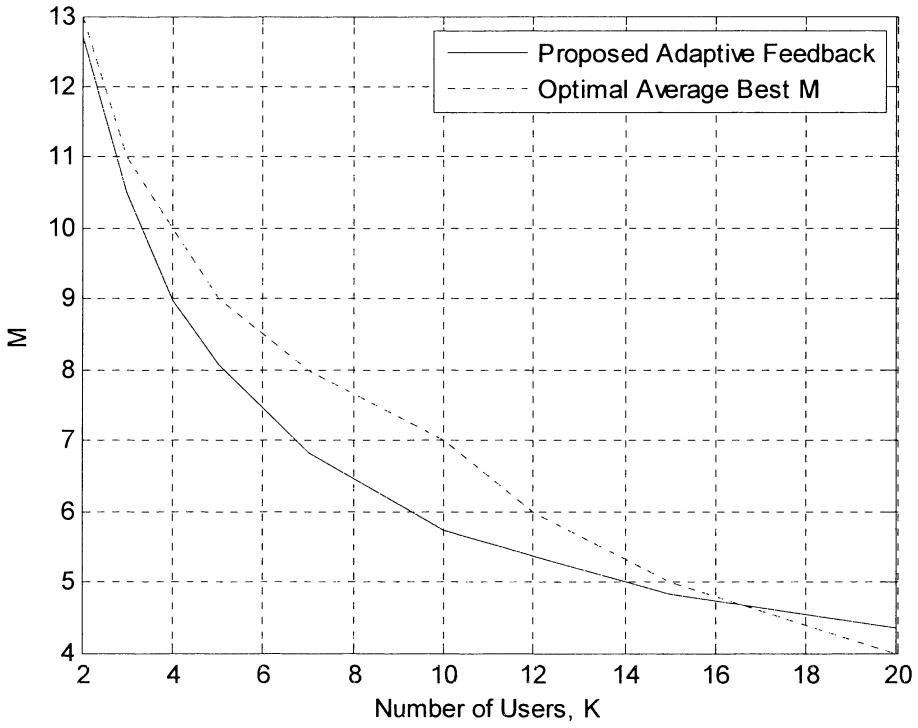


Figure 6.8 Comparison of parameter *M*

Figure 6.8 demonstrates that the average value of the parameter *M* of the proposed algorithm behaves like the optimally selected parameter *M* of the average best *M*. The proposed adaptive feedback scheme adjusts M to a large value (approximately *M=13*) when the number of users is low to ensure that the feedback outage is minimised. On the other hand, it reduces the value of *M* when the number of users becomes high to minimise the probability that two or more users are reporting CSI on the same subchannel. The result in Figure 6.8 validates the similar performance between the proposed adaptive feedback and the average best *M* feedback with optimal *M* in Figure 6.7.

Figure 6.9 shows the adaptive parameter *M* of the proposed feedback scheme for *K =2* and *K =20*. It can be observed from Figure 6.8 that, for *K = 2*, the values of adaptive M are varying around *M = 13*, while, for *K =20*, the values are varying around *M = 4*. The variation in the parameter *M* is more significant for *K =20*. This is because more users introduce more variation to the system.

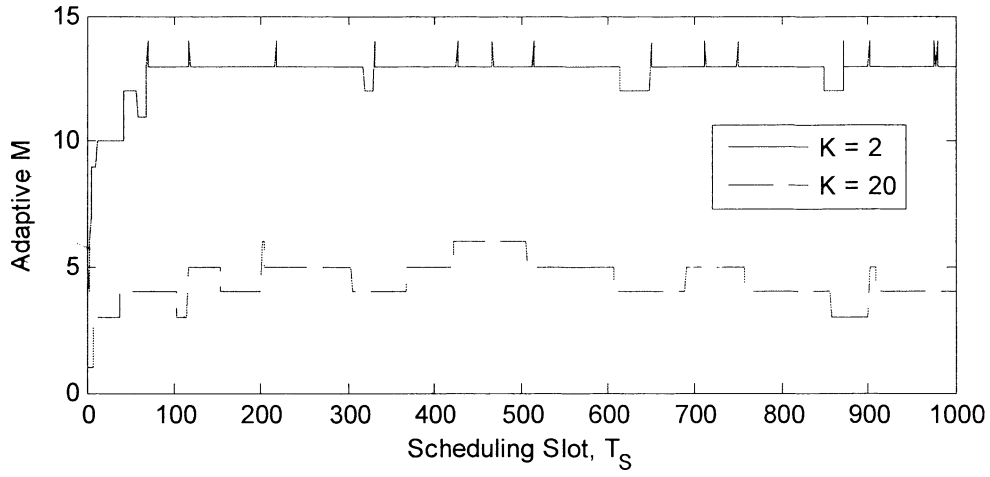


Figure 6.9 Snapshot of adaptive parameter M

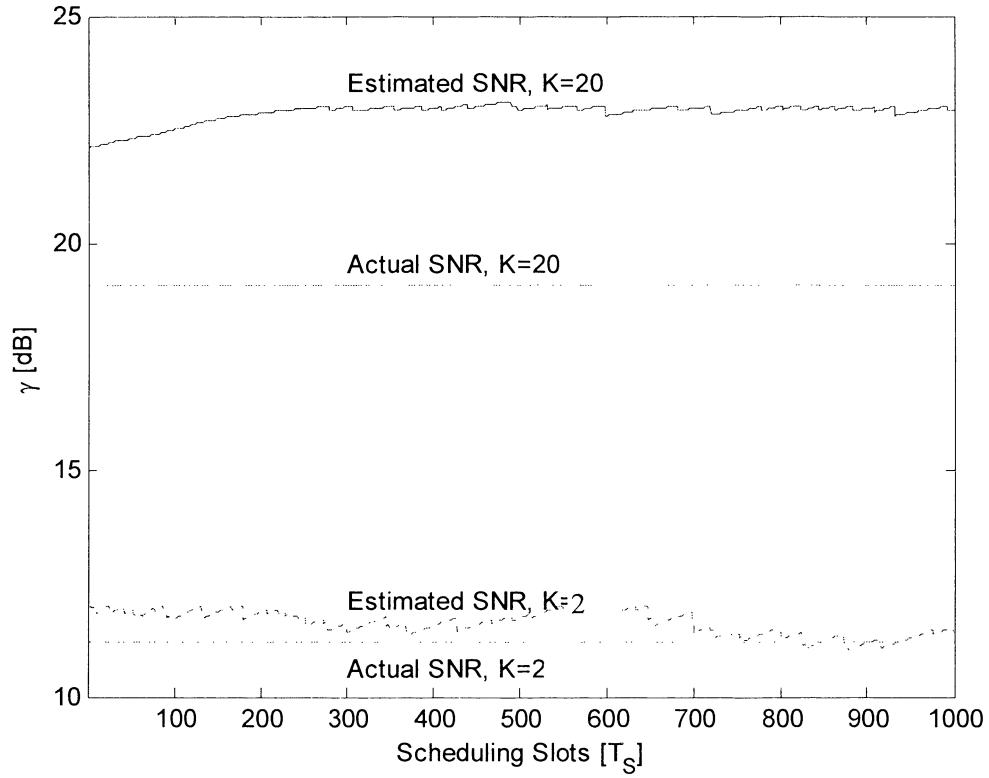


Figure 6.10 Snapshot of SNR reconstruction from ACK/NACK

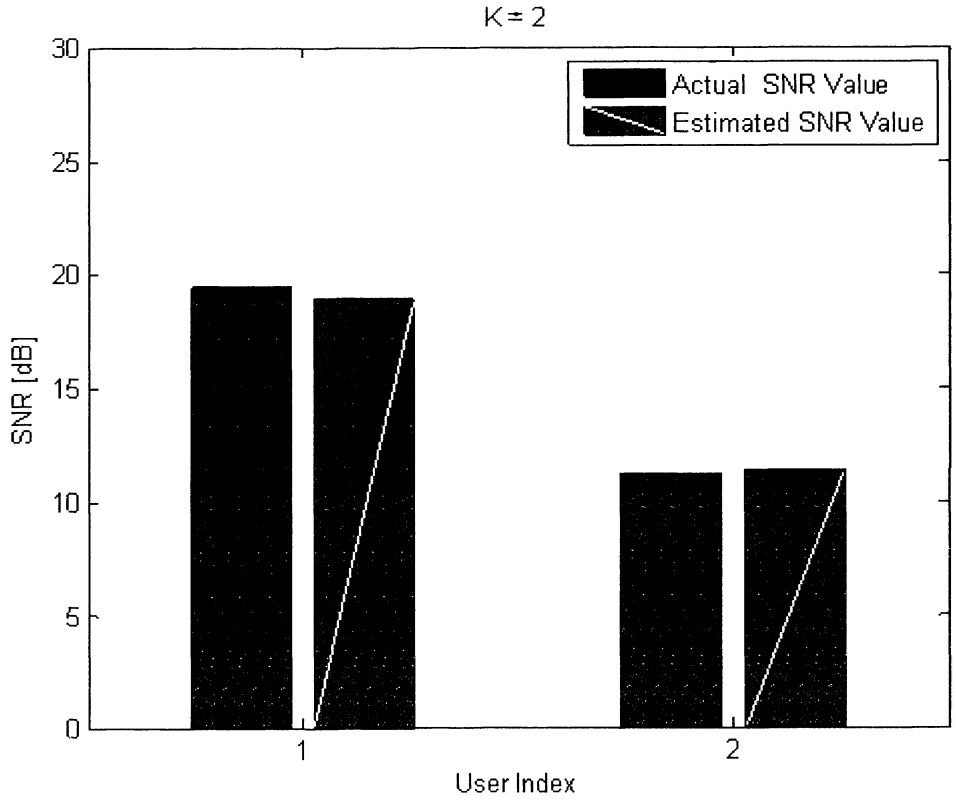


Figure 6.11 Average of SNR reconstructed from ACK/NACK vs. real SNR, $K = 2$

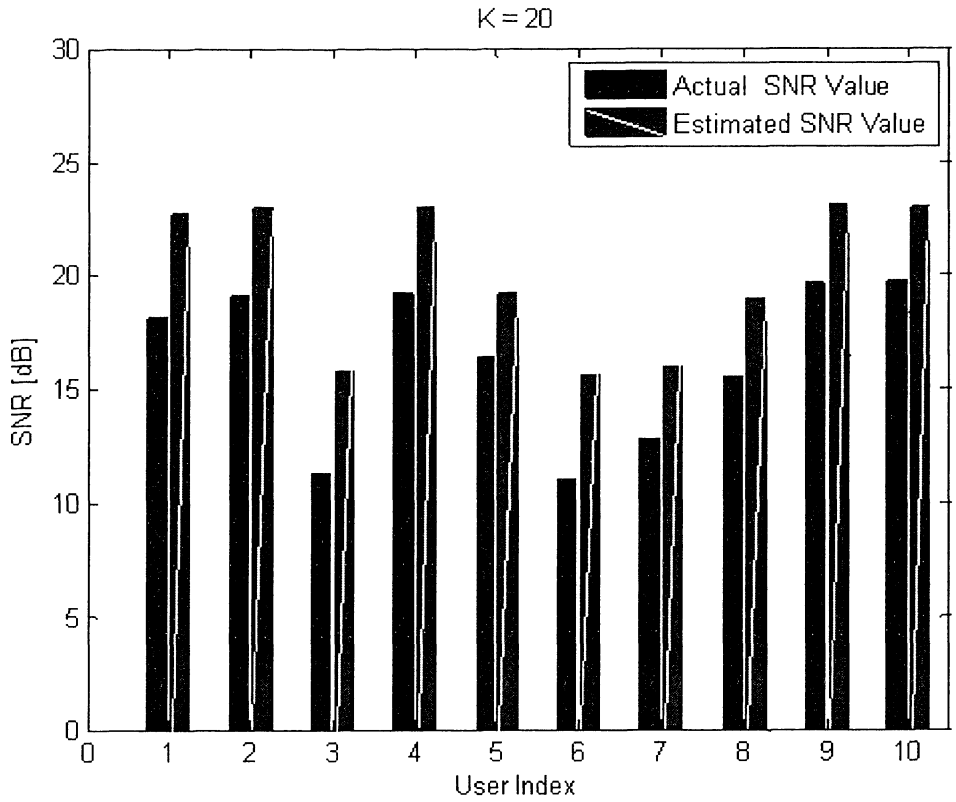


Figure 6.12 Average of SNR reconstructed from ACK/NACK vs. real SNR, $K = 20$

Figure 6.10, 6.11 and 6.12 shows the comparison of the estimated SNR extracted from ACK/NACK as mentioned in Section 6.3.2 and the real SNR value.

In Figure 6.12, only 10 users out of 20 users are shown for the clarity of presentation.

The results in Figure 6.10 to 6.11 indicate that mean SNR value for each user can be reconstructed from ACK/NACK messages. It can been seen from these figures that the reconstructed SNR for low number of users (i.e. $K = 2$) is closer to the real SNR than the estimation when the number of users is large (i.e. $K = 20$). This is due to the fact that this proposed technique can generate SNR values that capture the effect of multiuser diversity gain, which is desirable for our purpose to use the estimated SNR to determine the supportable MCS.

Figure 6.13 shows the effect of the adaptive M updating step sizes, ΔM_u and ΔM_d, on the performance of the proposed feedback scheme.

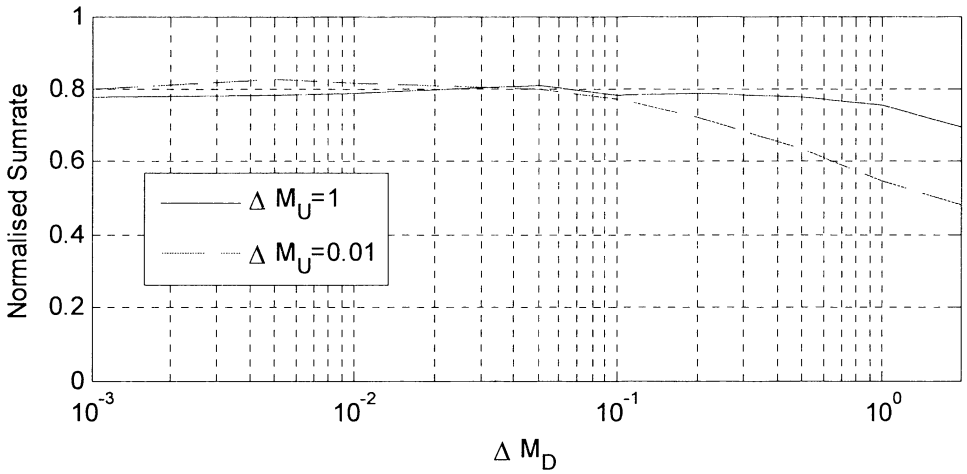


Figure 6.13 Effect of adaptive M updating step sizes

The results in Figure 6.13 demonstrate that the selection of the values of parameters ΔM_u and ΔM_d affects the performance of the proposed feedback scheme. In particular, these parameters should be selected such that $\Delta M_u >> \Delta M_d$. This is because number of allocated subchannels are much more likely to be less than the M value ($M_A < M$), especially when the number of users is moderate to high. In this

case, a comparatively large ΔM_d will result in continuously decreasing the adaptive-M values. It can be observed from Figure 6.13 that a wide range of $\Delta M_d \in [0.001 , 0.1]$ is compatible with $\Delta M_u = 1$.

It must be noted that the selection of the values of parameters ΔM_u and ΔM_d is required for the proposed adaptive feedback technique only at the beginning of the system deployment. On the contrary, the optimum M for adaptive best M and optimum feedback threshold γ_{th} for one-bit feedback schemes have to be continuously calculated and broadcasted to all users by the base station during the data transmission session.

6.6 Summary

In this chapter, a new adaptive feedback scheme for downlink OFDMA systems is proposed. The proposed scheme automatically adjusts the value of M to suit the number of users in the system. Simulation results demonstrate that, at K = 5, the proposed feedback scheme outperforms average best M technique with M = 5 by 20.8% and outperforms best M with M = 3 by 16.1%. The proposed technique achieves a similar performance as the average best M feedback with optimally selected M and the one-bit feedback algorithm with optimally selected feedback threshold without any requirement for the base station to calculate and broadcast feedback parameters during the data transmission session.

It is demonstrated that the supportable MCS of an OFDMA system with multiuser diversity gain can be determined by using SNR which is reconstructed from ACK/NACK messages.

Chapter 7

Opportunistic Contention-based Feedback for OFDMA

In the previous chapters, several feedback reductions techniques have been evaluated. In Chapters 3 to 5, the feedback threshold is used to limit the number of users sending feedback. Although the number of users sending feedback is significantly reduced by the use of the feedback threshold, there is a need for a dedicated feedback channel for each user which means that the resources allocated to carrier feedback are the same as the full feedback. This implies that the feedback channel must be sized to suit the maximum number of users that the system is designed to support. This may lead to an inefficient use of feedback resource as users are expected to send feedback when their received SNR is superior only. To improve the efficiency for the utilisation of the feedback resource, a contention-based feedback channel where all users attempt to send feedback on relatively small number of feedback channels can be deployed.

This chapter is organised as follows. Section 7.1 provides an overview of contention-based feedback techniques presented in the literature. A new feedback reduction technique for OFDMA systems that limit the feedback resources required is proposed in Section 7.2. Section 7.3 provides the simulation disceiption and associated simulation parameters. Simulation results are provided and discussed in Section 7.4. Section 7.5 draws a conclusion to this chapter.

7.1 Overview of Contention-based Feedback Techniques

Various approaches have been proposed to minimise the resource allocated for feedback using fixed contention-based channels. In [53], a feedback reduction protocol was designed for single carrier systems where all users compete to send feedback via a shared contention-based feedback channel. The idea of the contention-based feedback in [53] is to have a number of minislots in the uplink feedback channel where users compete to send feedback messages. The users

whose SNR exceeds a threshold compete for one of the minislots with an optimized, fixed and predetermined probability. A frame structure of a system implementing contention based feedback is shown in Figure 7.1.

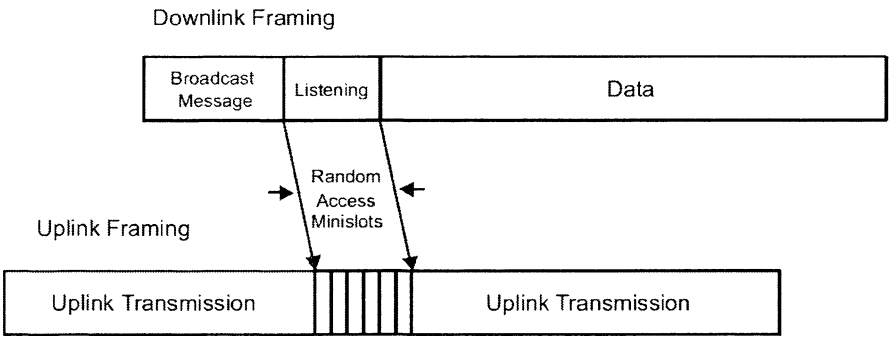


Figure 7.1 Frame structure of Tang's contention-based feedback protocol [53]

In Figure 7.1, at the beginning of each downlink frame, the base station broadcasts messages which contain relevant information to the feedback procedure to notify of the feedback period. The message is followed by S feedback minislots in the uplink direction. These minislots are random access contention-based channels. The number of minislots, S, is determined at the beginning of the system deployment and fixed throughout the operation. Users will attempt to access a feedback minislot when their SNR exceeds a predefined threshold that is dynamically updated to suit the system conditions (i.e. number of users).

The same idea was extended for OFDMA systems in [53]. The idea of the use of contention-based feedback scheme was extended in [54] by dividing users into multiple groups each of which has a separate set of feedback minislots. In [55], an OFDMA resource allocation with multiple-threshold contention-based feedback protocol was proposed. In [56], a technique to implement contention-based feedback protocol with multiple thresholds in which users compete for a set of feedback minislots based on their SNR was introduced.

7.2 Proposed Contention-based Feedback Protocol

In this section, a new protocol to reduce the amount of the feedback resources needed for OFDMA systems with mixed realtime (RT) and non-realtime (NRT) traffic is proposed. For each RT user, a dedicated CSI feedback minislot is

assigned. This is to ensure that the CSI feedback for RT users is always sent to the scheduler (base station, BS) successfully. On the other hand, for NRT users, the feedback resources are obtained by random access minislots and the number of minislots is set as a constant which is determined during the system deployment phase. The assumption that the number of feedback minislots are fixed is similar to the assumption in [53][54][56], in which all users compete for a minislot on *each* subchannel[9]. The protocol proposed in this chapter allows all NRT users to compete for a minislot on *any* subchannel to send CSI.

The parameters used in this chapter are defined first.

N = number of subchannels

N_{FB} = number of CSI feedback subchannels ($N_{FB} = N + K_{RT}$)

K_{RT} = number of RT users = number of dedicated feedback minislots for RT users

K = NRT users ($K_{total} = K + K_{RT}$)

K_{FB} = number of NRT users participating in minislot competition.

S = contention-based minislots per subchannel for NRT users only

$S \cdot N$ = Number of contention-based feedback minislots for NRT users

M = number of subchannels in which CSI will be reported

γ_{th} = feedback threshold

N_{th} = number of subchannels whose SNR must be above γ_{th} to send feedback

Each subchannel contains S *contention-based minislots*, which implies $S \times N$ contention-based minislots in total. All NRT users can compete for the entire $S \times N$ minislots. Contention may occur on these $S \times N$ minislots only. The message to be sent by the user on the contention minislot is the user ID.

[9] In this chapter, we assume that the smallest resource element in OFDMA is subchannel which is a group of subcarriers over a scheduling slot similar to resource block in LTE . .

After the *S* contention-based minislot, there is a *non-contention-based CSI feedback channel* which is allocated to a user who successfully competes for a contention minislot to send CSI feedback (see Figure 7.2).

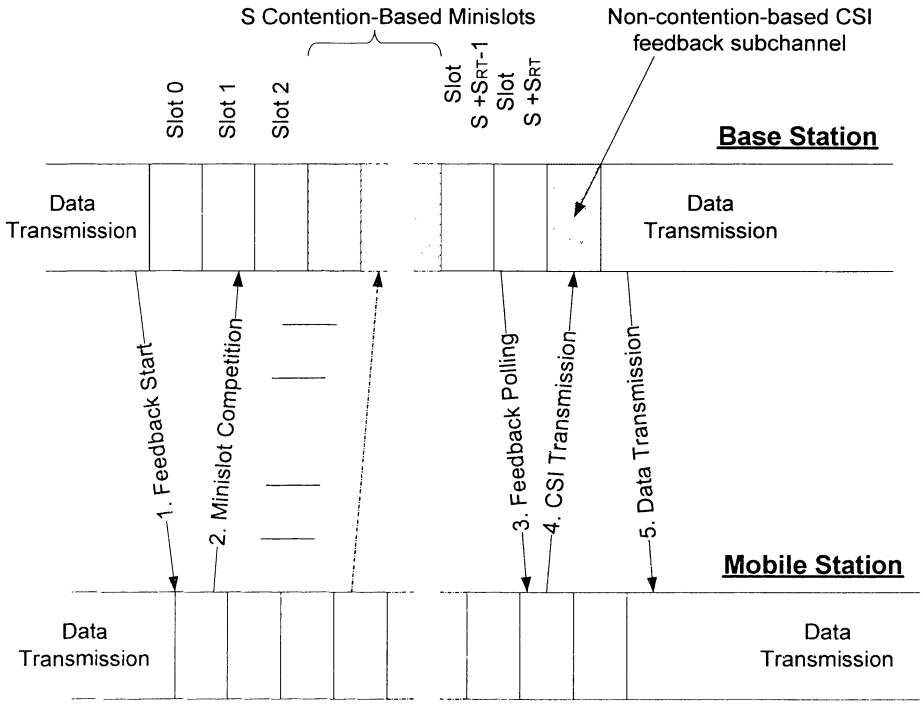


Figure 7.2 The proposed feedback procedure.

Each CSI feedback subchannel can carry only one CSI message from one user: the number of CSI feedback channels is equal to N_{FB} ($N_{FB} = N + K_{RT}$). Therefore, if the number of the NRT users who successfully obtain a contention minislot exceeds the number of CSI feedback channels, the BS randomly selects N users out of the successful users to send CSI feedback.

The proposed protocol (for one subchannel) is demonstrated in Figure 7.2.The steps involved in the proposed protocol are explained below:

Step 1: Feedback Start

The BS broadcasts a feedback message containing information regarding the opportunistic feedback to all MSs. This message includes N_{th} and γ_{th} parameters that will be discussed in Step 2.

Step 2: Minislot Competition

When the MSs have received the broadcast message, they start competing for feedback minislots. To optimally allow users to attempt to compete for a feedback minislot, a new feedback selection criterion is introduced. Users having SNR above a predefined threshold on a number of subchannels are allowed to compete for a minislot to send feedback. The parameter β_k is defined as the number of subchannels having SNR above the threshold for user k, which can be expressed as:

$$\beta_k = \sum_{n=1}^{N} U(\gamma_{kn} - \gamma_{th}) \tag{7.1}$$

where γ_{th} is the received SNR threshold, $U(x) = 1$ when $x > 0$, and $U(x) = 0$ when $x \leq 0$. User k is allowed to attempt to compete for a minislot if and only if $\beta_k > N_{th}$ where N_{th} is the minimum number of subchannels on which signal quality is above the threshold γ_{th}. The parameter N_{th} is set by the BS and broadcasted to all users. The parameter β_k is known to the MS only. Figure 7.2 shows an example in which a user has $\beta_k = 5$ subchannels that satisfy the criteria $\gamma_{n,k} > \gamma_{th}$.

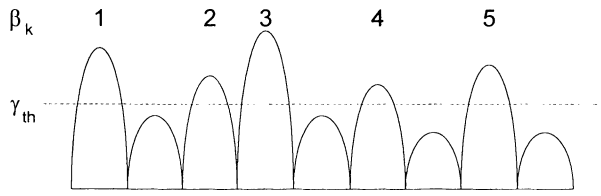


Figure 7.3 Proposed selection criterion

Users who successfully compete for a minislot send a message containing their User ID to the BS. If two or more users attempt to send feedback on the same feedback minislot, a collision occurs and the feedback is assumed to be irrecoverable.

Note that RT users do not participate in the minislot competition since dedicated CSI feedback channels will be allocated to RT users as discussed in Step 3.

Step 3: Feedback Polling

After the contention phase is finished, the BS knows the number of users who successfully compete for the minislots and their user IDs. The BS then selects N out

of K_{FB} NRT users whose received SNR met the criterion in Equation 7.1 and broadcasts a polling message to all users to identify which users are allowed to send CSI on each subchannel. Note that all RT users are always included in the feedback polling message.

Step 4: CSI Feedback Transmission

A user whose user ID appears in the polling message of a subchannel sends a CSI message to the BS. In this work, the best-M feedback technique [47] is used to reduce the feedback load for both RT and NRT users. Each user sends a CSI message containing the SNR levels of M highest subchannels.

Step 5: Data Transmission

Once the BS received CSI messages from all N NRT users indicating their best M subchannels, it performs subchannel allocation based on the received CSI. A simple subchannel allocation where each subchannel is allocated to the user having the highest CSI on this subchannel is used. The achievable average capacity for subchannel n thus can be given by:

$$C_n = E\left\{\log_2(1+\max_{k'} \gamma_{k'n_{k'}})\right\} \tag{7.2}$$

where k' represents the set of users who successfully send CSI feedback, and n_k' represents the set of index of M subchannels having highest SNR of the k^{th} user. $E\{x\}$ denotes the mean value of random variable x.

To better explain the contribution of our work, let us consider an OFDMA system supporting K users, N subchannels, and SxN feedback minislots (only NRT users are considered as RT users always have dedicated feedback channels and do not participate in contention phase). With other contention-based feedback protocols such as in [54][57], S feedback minislots will be assigned for each subchannel. Since the number of minislots per subchannel is small, the probability of unsuccessful access is high resulting in collisions. However, in the proposed protocol, all SxN minislots are available for NRT users to compete to send feedback

based on a new selection criterion to optimally minimise the number of users sending feedback as discussed in Section 7.2 Step 2.

7.3 Simulation Description

Simulations were performed using MATLAB-based Monte Carlo simulations using an average SNR for all users fixed at 6 *dB*, $M = 5$, and $N = 32$, unless mentioned otherwise. Simulation parameters are summarised in Table 7.1. The defaults values given in Table 7.1 are values that are found to be optimal and used in simulation if not indicated otherwise in the respective section.

Table 7.1 Simulation parameters

Descriptions	Values
Average SNR	Same for all users at 6 dB
Users requirements	Homogeneous (same QoS requirement)
Number of subchannels	32
Scheduling policy	Maximum throughput
Scheduling slot	1ms
Channel model	Uncorrelated Rayleigh
CSI report	Best M technique
Supportable data rate	Shannon's formula
M	1 to 7 (5 is default value)
S	1 to 10 (5 is default value)
N_{th}	1 to 8 (1 is default value)
γ_{th}	Optimally calculated for different number of NRT users

The downlink data transmission of an OFDMA system having N subchannels is considered. The total transmit power is a constant equal to P_t. A frequency-selective

Rayleigh fading channel is considered in this work. It is assumed that the channel gains for all users are independent and identically distributed (i.i.d.) for all subchannels. The received signal-to-noise ratio (SNR) of user k on subchannel n is defined as $\gamma_{kn} = |h_{kn}|^2/\sigma^2$ where σ^2 denotes the variance of the zero-mean additive white Gaussian noise (assumed $\sigma^2 = 1$). For user k on subchannel n, the received SNR γ_{kn} is assumed to be perfectly known at the respective user, and the base station may know the SNR by CSI sent from the users only. All users are assumed to have homogeneous SNR. The feedback links are assumed to be delay-free and error-free.

7.4 Simulation Results

In this section, the focus is on the performance evaluation of the proposed scheme and a comparison with the Tang's contention-based feedback protocol in [54]. To better demonstrate the performance of the proposed protocol for RT and NRT users, the simulation results for RT are discussed in Section 7.4.1 while the results for NRT are discussed in Section 7.4.2.

7.4.1 Performance of RT Users

With the proposed scheme, BS always allows RT users to send CSI with no collision. Once BS received the CSI from RT users, it provides a higher priority for RT users in the subchannel allocation. This means each RT user will be allocated its best subchannel with higher probability compared to a NRT user. Capacity loss may occur when different RT users have the same subchannels with highest SNR. However, the probability of this event is very low and thus the loss can be neglected.

The proposed scheme is compared with a simple subchannel allocation where a fixed subchannel is dedicated for a RT user.

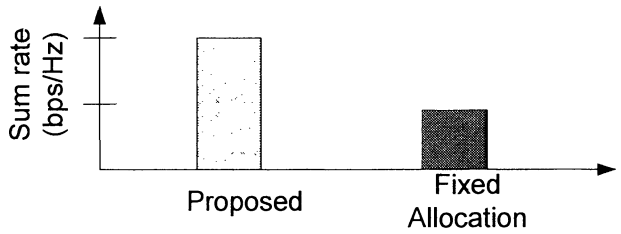


Figure 7.4 Achievable average sum-rate RT users

Figure 7.4 shows the achievable sum rate capacity of the proposed scheme and the fixed allocation. The proposed scheme can achieve much higher performance (approximately double) than the fixed allocation for RT users. This is because the proposed scheme is more likely to allocate to a user the subchannel on which it has the highest SNR, thus benefiting from frequency diversity. However, this performance gain comes with an increase in feedback load.

7.4.2 Performance of NRT Users

In this section, the performance of the proposed scheme for NRT users is studied. Figure 7.5 and 7.6 show the average sum-rate versus the number of NRT users K when the number of minislots per subchannel S is fixed at 5 and 10, respectively.

Figure 7.5 demonstrates that the performance of the proposed protocol is close to the full CSI feedback case for a moderate number of users (20-50). Overall, the proposed scheme significantly outperforms the scheme in [54] over a wide range of users. The sum rate improvement of the proposed technique suggests that users should be allowed to compete for all available contention minislots.

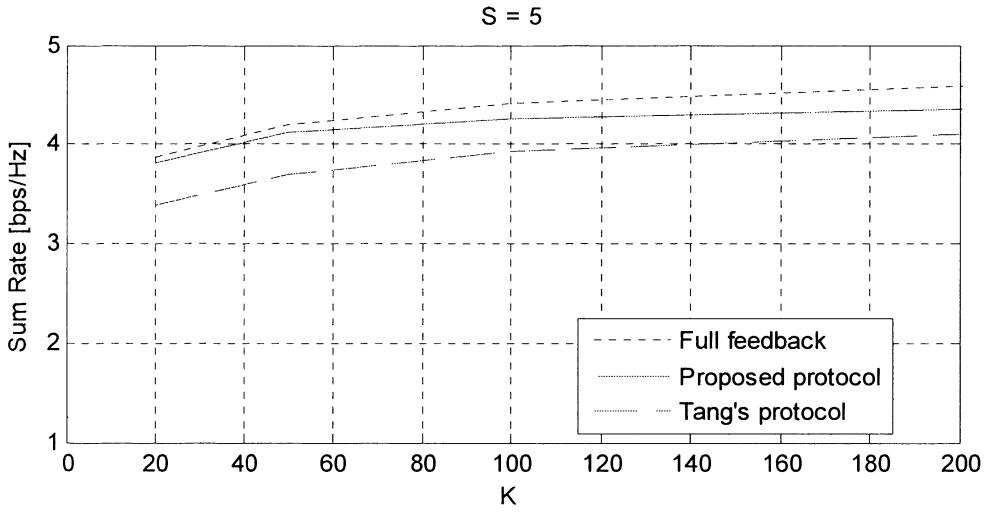


Figure 7.5 The achievable average sum-rate versus number of users, S = 5

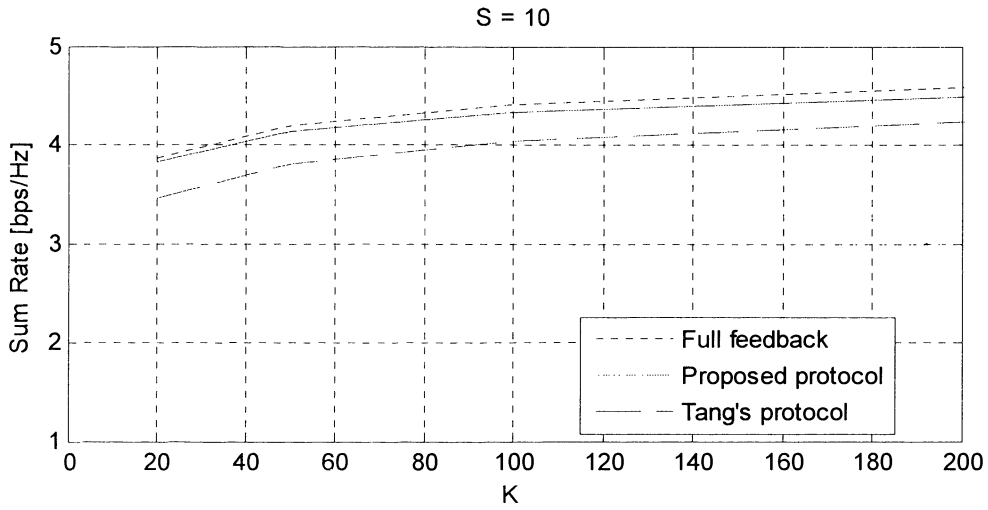


Figure 7.6 The achievable average sum-rate versus number of users, S = 10

Table 7.2 Sum rate comparison for various K when S=5.

Number of NRT users, K	**20**	**50**	**100**	**200**	**Average**
Proposed technique compared with full-CSI	99.8%	98.1%	96.5%	95.0%	97.1%
Tang's technique compared with full-CSI	87.7%	88.2%	89.3%	89.4%	88.7%
% Improvement	11.1%	9.8%	7.2%	5.5%	8.4%

There are four parameters involved in optimising the proposed scheme namely γ_{th}, *Nth*, *S*, and *M*. The values of *S* and *M* need to be determined in the system deployment phase. The parameters that can be varying and optimised dynamically for each system condition (i.e. number of users) are γ_{th}, and N_{th},.

7.4.2.1 Optimal γ_{th}, and N_{th}

This section evaluates the effect of the parameter N_{th} and γ_{th} on the sum rate performance.

Figure 7.7 shows the average achievable sum rate versus γ_{th}, for N_{th} = 1, 2 and 3. The total number of NRT users, *K*, is set at 500 and the number of contention minislots per subchannel, *S*, is 5.

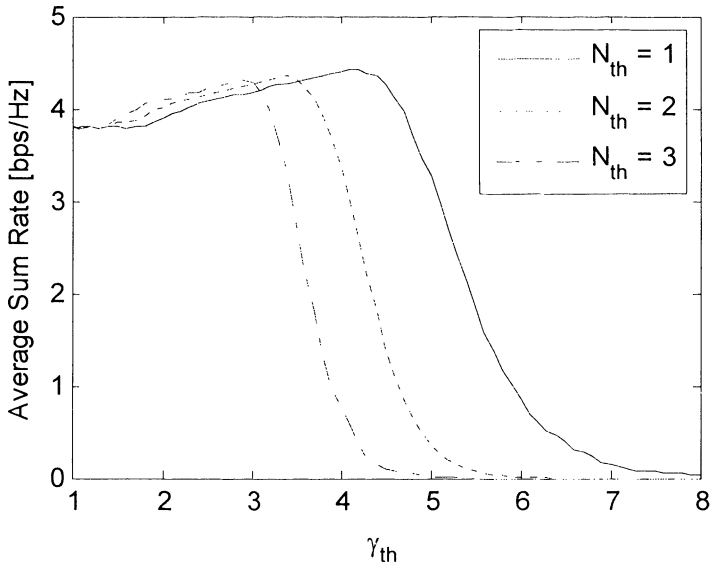


Figure 7.7 Effect of γ_{th}, on the sum rate

Figure 7.7 indicates that there is an optimum γ_{th} that maximises the achievable sum-rate for each N_{th}. It can be observed from Figure 7.5 that N_{th} = 1 results in the highest achievable capacity. This is because having N_{th} = 1 will allow only users whose SNR is relatively high to attempt to send CSI. Since the BS picks *N* out of K_{FB} feedback users who successfully competed for a minislot to send CSI feedback, thus enabling data transmission from users who have relatively high SNR

7.4.2.2 Optimal M

This section evaluates the effect of the selection of the number of subchannels in which CSI should be reported (M) on the sum rate performance.

Figure 7.8 shows the sum rate of the proposed technique when the parameter M increases from 1 to 7 when K = 100 and S =5. The sum rate of full CSI feedback and Tang's contention based technique is plotted for comparison.

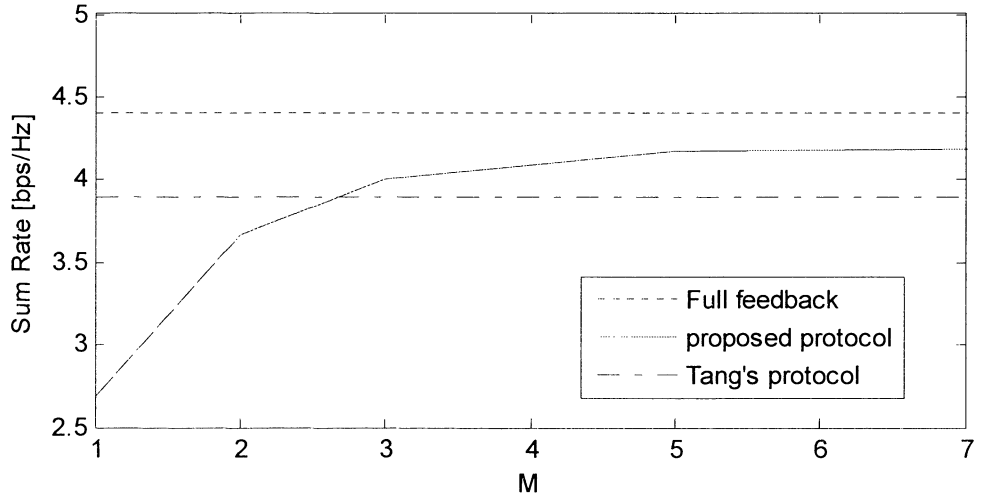


Figure 7.8 Sum rate when M changes

From the simulation result in Figure 7.8, the proposed protocol requires the transmission of at least the CSI of three subchannels ($M \geq 3$) to overcome the achievable sum rate of Tang's protocol. It can be observed that the average sum rate increases with an increase of M. However, the increasing sum rate saturates at M=5. Therefore, M = 5 is considered as an optimal.

7.4.2.3 Optimal S

This section determines the effect of parameter S on the sum rate of the system.

Figure 7.9 shows the achievable sum rate of the proposed feedback technique for various S in comparison with Tang's technique. In this simulation, K is set at 100, and M is set at 5. The sum rate of full CSI feedback is plotted as a performance benchmark.

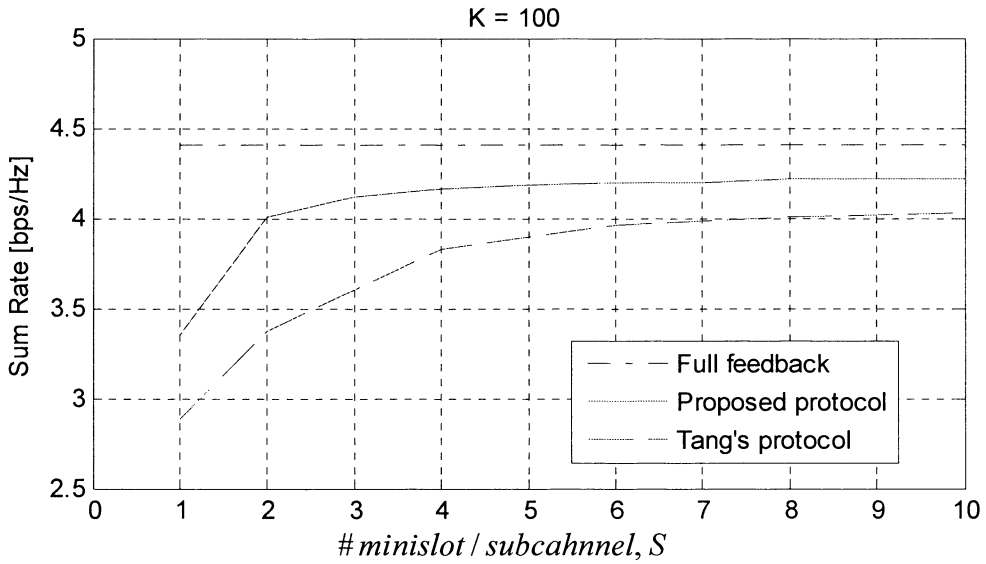


Figure 7.9 Average sum-rate versus the number of minislots per subchannel (*M*).

Table 7.3 Sum rate comparison for various S when K = 100.

Number of contention minislots per subchannel, S	**1**	**2**	**5**	**8**	**10**	**Average**
Proposed technique compared with full-CSI	76.1%	91.1%	95.0%	95.8%	95.8%	90.8%
Tang's technique compared with full-CSI	65.5%	76.6%	88.5%	91.1%	91.6%	82.7%
% Improvement	10.6%	14.5%	6.5%	4.7%	4.2%	8.1%

Overall, the proposed scheme significantly outperforms Tang's protocol in [54], especially when the number of minislots per subchannel is small. The result indicates that, with a small number of minislot such as when $S = 2$, our proposed scheme can achieve a high proportion of the average sum-rate (4.1 bps/Hz out of 4.4 bps/Hz i.e. 91.1%) while the protocol in [54] causes a large capacity loss in the achievable sum rate (only 76.6% of the full CSI can be achieved). This is corresponding to an improvement 14.5% from Tang's protocol with the same amount of feedback resources required. The result indicates that the proposed

feedback technique can achieve more than 90% of the full CSI sum rate when the feedback minislots are limited to only 2 minislots per subchannel.

Furthermore, the result in Figure 7.9 shows that the proposed technique requires 2 minislots/subchannel to achieve 4 bps/Hz while Tang's technique requires 6 minislots/subchannel. This result indicates that the proposed technique requires less feedback resource during the feedback competition phase than Tang's technique.

7.4.2.4 Optimum γ_{th}

This section evaluate the effect of the feedback threshold, γ_{th} on thesum rate for S = 1, 2 and 10 when K = 1 and N_{th} = 1.

Figure 7.10 shows the achievable sum rate of the proposed feedback technique for various S in comparison with Tang's technique. The sum rate of full CSI feedback is plotted as a performance benchmark.

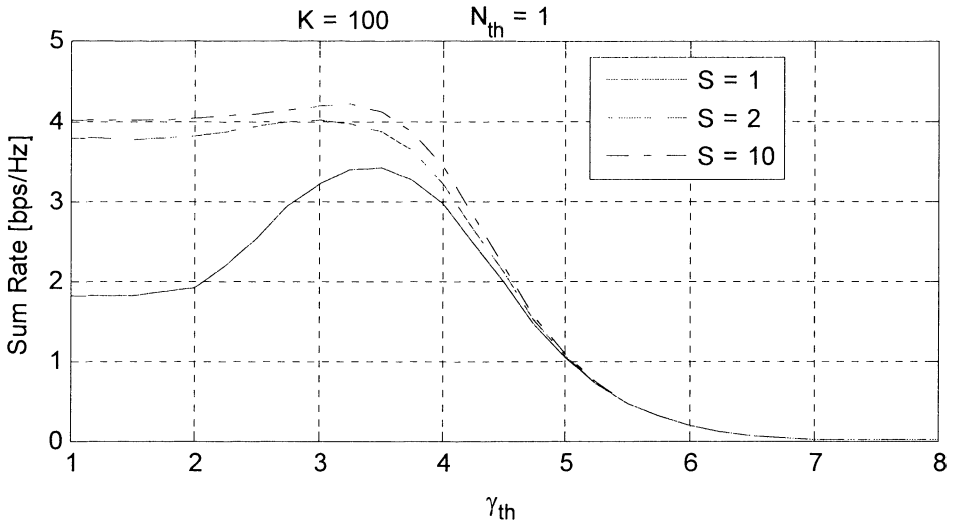


Figure 7.10 Effect of the parameter γ_{th}.

Figure 7.10 demonstrates that there exists an optimum γ_{th} for S =*1, 2* and *10*. Furthermore, it is shown that the choice of γ_{th} is more critical for a small S. This is because the value of the parameter γ_{th} plays a critical role in controlling the number of users that are allowed to compete for a minislot. It can be observed that the optimal γ_{th} is around 3.5 for S = 1. The sum rate decreases rapidly when γ_{th} is smaller than 3.5 because an excessive number of users attempt to compete for the

minislots. However, in cases with large S (S = 2, and S =10), having a small γ_{th} causes a negligible effect on the average sum-rate. This is because they have enough minislots to ensure a low collision rate when many users are attempting to send feedback. For S = 1, the sum rate decreases rapidly when an excessively large γ_{th} is employed because the probability that users have at least one subchannel having quality higher than 3.5 is low resulting in a situation where no user is attempting to send feedback. When γ_{th} is high ($\gamma_{th} > 4$) , the performance loss due to feedback outage is independent of the number of available minislots (S). .

7.5 Summary

In this chapter, a new feedback protocol for OFDMA systems with mixed RT and NRT traffic is proposed. For RT users, a dedicated CSI feedback channel is allocated. For NRT, the protocol is based on the contention-based feedback resource where all NRT users can attempt to send feedback via feedback minislots on any subchannels. A NRT user is allowed to request for a feedback minislot only if the number of subchannels whose SNR is above the threshold is greater than a predefined threshold value.

Simulation results demonstrate that, for RT case, assigning a dedicated CSI feedback channel for each RT user results in significant performance gain over allocating a fixed subchannel to transmit data.

For the NRT case, the simulation results show that around 8% improvement in sum rate can be achieved when NRT users are allowed to compete on SxN minislots on the entire subchannels when compared with when they are allowed to compete on S minislots on each subchannel. When the feedback minislots are very limited (such as S = 2), the proposed scheme can achieve an improvement of 14.5% compared with Tang's technique. The achievable sum rate of the proposed scheme is 95% close to the full CSI feedback case for K = 20 to 200.

Chapter 8
Conclusions and Future Work

8.1 Conclusions

Chapter 1 provides an introduction to the thesis. Furthermore, the problem statements and the aims of this work as well as research contributions are outlined.

Chapter 2 introduces some background knowledge of LTE network architecture, OFDMA technology, adaptive modulation techniques, and system modelling. Details of simulations are briefly explained in this chapter.

Chapter 3 attempts to answer the research question "Based on our current knowledge of feedback reductions techniques, can the feedback load be further reduced without scarifying the system performance?". In this chapter, a new feedback reduction technique that sends two or more bits for a CSI command per scheduling slot is presented and evaluated. It is demonstrated that the probability that the proposed feedback reduction technique with δ_{max} =3 (two bits required) can achieve efficient CSI report in order of 99% accuracy. Simulation results verify that the net throughput obtained by the proposed scheme with two-bit feedback outperforms that obtained by the one-bit feedback scheme by 25.3% on average. Approximately the same net throughput as Gesbert's selective feedback can be achieved with around 40% feedback reduction.

Chapter 4 deals with the research question of "Do multiple threshold schemes perform better than single threshold schemes?". In this chapter, the achievable sum rate of multiuser systems with multiple feedback thresholds is formulated. It is demonstrated that the system with two thresholds can achieve 98% of the full-CSI sum rate which is equivalent to around 8.5% improvement compared with the system with one threshold. Furthermore, suboptimal thresholds for system with two

thresholds are evaluated. The suboptimal threshold value of α_1 can be generated by applying 2 dB to the optimal threshold value α_0.

In the first part of Chapter 5, the research question: "How is it possible to develop a feedback technique(s) that can self-adjust the parameters associated with CSI reporting to suit the system conditions?" is addressed by proposing a new feedback reduction technique where the feedback threshold can be automatically adjusted is proposed. The proposed technique can self-tune the feedback threshold to suit the number of users.

The second part of Chapter 5, deals with the research question: "Provided that a well-known scheduling technique creates significant delays in the scheduling process, it is possible to mitigate the delays without any need for additional feedback resources?". In this chapter, a new hybrid scheduling algorithm with reduced feedback, named HPF, which can decrease the scheduling delay is proposed. HPF behaves like PFS when one or more users send(s) feedback. If none of the users send CSI feedback the resources will be allocated to the user having the largest delay (similar to FIFO). Simulation results demonstrate that the proposed algorithm can reduce P_{ex} by 33.3%, and τ_{max} by 45% on average compared with PFS. HPF outperforms EXPR in the heavily loaded network by more than 30% in term of achievable sum rate. The implementation of HPF does not involve any hardware upgrading.

Chapter 6 addresses the research question "Provided that the best-M scheme when CSI's of best M subchannels out of all N subchannels are reported causes significant capacity loss when the number of users is small, is it possible to improve the algorithm so that it works well when the number of users is small?". In this chapter, a new adaptive feedback scheme for downlink OFDMA systems is proposed. The proposed scheme automatically adjusts the value of M to suit the number of user in the system. Simulation results demonstrate that, at K = 5, the proposed feedback scheme outperforms average best M technique with M = 5 by 20.8% and outperforms best M with M = 3 by 16.1%. The proposed technique can self-tune the feedback threshold for various numbers of users to achieve a similar

performance to the average best M feedback with optimally selected M and to the one-bit feedback algorithm with optimally selected feedback threshold. Furthermore, it is demonstrated that the supportable MCS of an OFDMA system with multiuser diversity gain can be determined by using SNR which is reconstructed from ACK/NACK messages.

Chapter 7, deals with the research question "Based on our current knowledge of CSI transmission on the uplink, can a more efficient media access control scheme for CSI reporting be developed?". In this chapter, a new feedback protocol for OFDMA systems with mixed RT and NRT traffic is proposed. Simulation results demonstrate that, for RT case, assigning a dedicated CSI feedback channel for each RT user results in significant performance gain over allocating a fixed subchannel to transmit data. For the NRT case, the simulation results show that around 8% improvement in sum rate can be achieved when NRT users are allowed to compete on S_xN minislots on the entire subchannels when compared with when they are allowed to compete on S minislots on each subchannel. When the feedback minislots are very limited (such as $S = 2$), the proposed scheme can achieve an improvement of 14.5% compared with Tang's technique. The achievable sum rate of the proposed scheme is 95% close to the full CSI feedback case for $K = 20$ to 200.

8.2 Future Work

This thesis has investigated a number of link adaptation and feedback reduction techniques for modern wireless networks. Simulation has been set to a reasonable complexity so that the PhD study could be completed within the allowed time frame. For the future study based on the content of this thesis, several aspects are suggested below

- All of the simulations performed in this thesis were based on a multiuser scenario in a single-cell OFDMA system. This simulation condition is adequate for evaluating the performance of link adaptation technique with reduced CSI in terms of achievable sum-rate. This simulation can be

extended to a more realistic scenario by adding more cells. In this case, the interference between cells would become an issue and coordination between cells would be applicable.

- The technology in this field has developed rapidly. In recent months, the specifications for 4G have been defined. In 4G system, the system performance is recommended to be enhanced by the use of relay network. Future work could extend the work in this thesis to develop a new feedback reduction technique in relay networks.

- In this thesis, it was assumed that the feedback channel is perfect i.e. no CSI error and delay. This assumption is not always valid in a practical situation where feedback may be erroneous. All work in this thesis can be extended to a situation where feedback channel is imperfect taking into account erroneous bits and delays.

- In this thesis, it was assumed that user buffers are always full containing packets to be sent on the downlink which is considered as a worst case scenario. One possible extension of this would be to introduce more realistic traffic flows and have different classes of user services requirements.

References

[1] 3GPP, RP 040125, "Overview of 3GPP Release 99, Summary of all Release 99 Features", March 2004.

[2] 3GPP, TR25.306, v 6.9.0, "Technical Specification Group Radio Access Network, UE Radio Access capabilities (Release 5)", Dec 2005.

[3] H.Holma and A. Toskala, HSDPA/HSUPA for UMTS, John Wiley & Sons, 2006, England.

[4] H.Holma and A. Toskala, LTE for UMTS, John Wiley & Sons, 2010, England

[5] 3GPP, RP 040461, "Proposed Study Item on Evolved UTRA and UTRAN", Dec 2004

[6] UTRA-UTRAN Long Term Evolution (LTE) and 3GPP System Architecture Evolution (SAE), http://www.3gpp.org/Highlights/LTE/LTE.htm, accessed 07/02/07.

[7] 3GPP, TR 25.913, "Requirements for Evolved UTRA and Evolved UTRAN", March 2006

[8] R. Knopp and P.A Humblet, "Information capacity and power control in single-cell multiuser communications", in Proc. IEEE ICC, Seattle, WA, Jun. 1995, pp. 331 – 335.

[9] P. Viswanath, D. N. C. Tse, and R. Laroia, "Opportunistic beamforming using dumb antennas," IEEE Trans. Inform. Theory, vol. 48, no. 6, pp. 1277–1294, Jun 2002.

[10] S. Sesia, I Toufik, M. Baker, LTE – The UMTS Long Term Evolution: From Theory to Practice, John Wiley & Sons, 2009

[11] F. Khan, LTE for 4G Mobile Broadband: Air Interface Technologies and Performance, Cambridge University Press, New York, 2009.

[12] S.T. Chung and A.J. Goldsmith, "Degrees of Freedom in Adaptive Modulation: A Unified View," *IEEE Trans. Comm.*, vol. 49, no. 9, pp. 1561-1571, Sept. 2001.

[13] J. F. Hayes, "Adaptive feedback communications." IEEE Trans. Commun. Technol., vol. COM-16, pp. 29-34, Feb. 1968

[14] A. Goldsmith, Wireless Communications, Cambridge University Press, New York, 2005.

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

[16] L. Yang, M. Kang, and M.-S. Alouini, "On the capacity-fairness tradeoff in multiuser diversity systems," IEEE Trans. Veh. Technol., vol. 56, no. 4, pp. 1901–1907, July 2007.

[17] K. I. Pedersen, G. Monghal, I. Z. Kov´acs, et al., "Frequency domain scheduling for OFDMA with limited and noisy channel feedback," in Proceedings of the 66th IEEE Vehicular Technology Conference (VTC '07), pp. 1792–1796, Baltimore, Md, USA, September-October 2007.

[18] H.Holma and A. Toskala, WCDMA for UMTS, John Wiley & Sons, 2002, England.

[19] H. Kim, and Y. Han, "An Opportunistic Channel Quality Feedback Scheme for Proportional Fair Scheduling" IEEE Commun. Letters, vol 11 no 6 June 2007.

[20] D. Gesbert and M.-S. Alouini, "How much feedback is multi-user diversity really worth?" in Proc. IEEE Int. Conf. on Communications (ICC '04) , Paris, France, pp. 234–238, Jun 2004.

[21] S. Sanayei and A. Nosratinia, "Opportunistic downlink transmission with limited feedback," IEEE Trans. Inform. Theory, vol. 53, no. 11, pp.4363–4371, Nov 2007.

[22] O. Somekh, A. M. Haimovich, and Y. Bar-Ness, "Sum rate analysis of downlink channels with 1-bit feedback," IEEE Commun. Letters, vol. 11, no.2, pp. 137-139, Feb 2007.

[23] H. Kim, and Y. Han, "An Opportunistic Channel Quality Feedback Scheme for Proportional Fair Scheduling" IEEE Commun. Letters, vol 11 no 6 June 2007. Communications Letters, IEEE Volume 11, Issue 6, pp. 501 – 503, Jun 2F007.

[24] F. Floren, O. Edfors, and B-A Molin, "The effect of feedback quantization on the throughput of a multiuser diversity scheme," in Proc. IEEE Global Telecommunications Conference, Sanfrancisco,CA, Dec. 2003.

[25] B. Holter,M.-S. Alouini, G. E. Øien, and H.-C. Yang, "Multiuser switched diversity transmission," in Proc. IEEE VTC, Los Angeles, CA, Sep. 2004, pp. 2038–2043.

[26] V. Hassel, M.-S. Alouini, D. Gesbert, G.E Oien, G.E., "Exploiting multiuser diversity using multiple feedback thresholds", Vehicular Technology Conference, Stockholm, Sweden, pp. 1302 – 1306, 30 May-1 June 2005.

[27] H. Nam and M. S. Alouini, "Multiuser Switched Diversity Scheduling Systems with Per-User Threshold," IEEE Trans. Commun, vol. 58, no. 5, pp. 1321–1326, May 2010.

[28] J. So and J. M. Cioffi, "Capacity and fairness in multiuser diversity systems with opportunistic feedback," IEEE Commun. Lett., vol. 12, no. 9, pp. 648-650, Sep. 2008.

[29] R. Elliott, "A measure of fairness of service for scheduling algorithms in multiuser systems," in Proc. IEEE CCECE, Winnipeg, MB, Canada, May 2002, pp. 1583–1588.

[30] M. Andrews, K. Kumaran, K. Ramanan, A. Stolyar, P. Whiting, and R. Vijayakumar, "Providing Quality of Service over a Shared Wireless Link," IEEE Communications Magazine, vol. 39, no. 2, pp. 150-154, 2001.

[31] S. Shakkotai and A. L. Stolyar, "Scheduling Algorithms for a Mixture of Real-Time and Non-Real-Time Data in HDR," 2000, pp. 1-12.

[32] A. K. F. Khattab and K. M. F. Elsayed, "Channel-quality dependent earliest deadline due fair scheduling schemes for wireless multimedia networks," in 7th ACM International Symposium on Modeling, Analysis and Simulation of Wireless and Mobile Systems, 2004, pp. 1-8.

[33] A. Gyasi-Agyei, and K., "Seong-Lyun Comparison of opportunistic scheduling policies in time-slotted AMC wireless networks," IEEE, ISWPC 2006, Phuket, Thailand, pp. 1 – 6, 16-18 Jan. 2006.

[34] D. Pareit, B. Lannoo, I. Moerman, and P. Demeester, "The History of WiMAX: A Complete Survey of the Evolution in Certification and Standardization for IEEE 802.16 and WiMAX," *Communications Surveys & Tutorials, IEEE* , vol.PP, no.99, pp.1-29, 0

[35] C. Y. Wong, R. S. Cheng, K. B. Letaief and R. D. Murch, "Multiuser OFDM with Adaptive Subcarrier, Bit and Power Allocation," IEEE J. on Selected Areas in Comm., Vol.17, No.10, pp.1747-1758, Oct. 1999.

[36] Z. Han and Y. Lee, "Opportunistic Scheduling with Partial Channel Information in OFDMA/FDD Systems", Proc. Of VTC'04 Fall, Vol.1, pp.511 – 514, Sept. 2004.

[37] J. Ko Y. Lee "Opportunistic Transmission with Partial Channel Information in Multi-User OFDM Wireless Systems", WCNC 2007. IEEE 11-15 March 2007 pp. 1318-1322.

[38] T. -S. Kang and H. -M. Kim, "Opportunistic feedback assisted scheduling and resoruce allocation in OFDMA systems," in Proc. IEEE ICCS, Vol. 1, pp. 1-5, Oct. 2006.

[39] C. Jieying, R. A. Berry, and M. L. Honig, "Large system performance of downlink OFDMA with limited feedback," in Proc. IEEE ISIT, 2006.

[40] S. Yoon, O. Somekh, O. Simeone, and Y. Bar-Ness, "A comparison of opportunistic transmission schemes with reduced channel information feedback in OFDMA downlink," in Proc. IEEE PIMRC, Sep. 2007, pp. 1–5..

[41] T. S. Kang and H. M. Kim, "Opportunistic feedback assisted scheduling and resource allocation in OFDMA systems," in Proc. IEEE ICCS, Oct. 2006, pp. 1–5.

[42] P. Svedman, S.K. Wilson, L. Cimini and B. Ottersten, "A Simplified feedback and scheduling scheme for OFDM", Proc. of VTC 2005, May 17-19, 2004, Milan

[43] IEEE 802.16e-2005, Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems Amendment for Physical and Medium Access Control Layers for Combined Fixed and Mobile Operation in Licensed Bands, IEEE, Feb. 2006.

[44] 3GPP TSG RAN v8.3.0, Evolved Universal Terrestrial Radio Access: Physical Layer Procedures (3GPP TS 36.213).

[45] "CQI design and its impact to DL performance," 3GPP TSG RAN WG1#48bis contribution R1-071682, March 2007.

[46] B. Chul Jung, T. Won Ban, W. Choi, andD. K. Sung; , "Capacity analysis of simple and opportunistic feedback schemes in OFDMA Systems," Communications and Information Technologies, 2007. ISCIT '07. International Symposium on , vol., no., pp.203-208, 17-19 Oct. 2007.

[47] N. Kolehmainen, J. Puttonen, P. Kela, T. Ristaniemi, T. Henttonen, and M. Moisio, "Channel quality indication reporting schemes for UTRAN long term evolution downlink," in Proc. 67th IEEE Vehicular Technology Conference 2008 (VTC2008-Spring), May 2008, pp. 2522 - 2526.

[48] S. Sanayei, A. Nosratinia and N. Aldhahir, "Opportunistic dynamic subchannel allocation in multiuser OFDM networks with limited feedback", Proc. of IEEE Info. Theory Workshop 2004 (ITW04), Oct. 2004, San Antonio, Texas.

[49] Somekh, O. Simeone, A.M. Haimovich and Y. Bar-Ness, "Sum-Rate Analysis of OFDM Downlink Channels with 1-bit Feedback per Subcarrier", Proc. of CISS' 06, pp. 306-311, Mar. 2006.

[50] K. Pedersen, T. Kolding, I. Kovacs, G. Monghal, F. Frederiksen, and P. Mogensen, "Performance Analysis of Simple Channel Feedback Schemes for a Practical OFDMA System," Vehicular Technology, IEEE Transactions on , vol.58, no.9, pp.5309-5314, Nov. 2009.

[51] Q. Liu, S. Zhou, and G. B. Giannakis, B "Cross-layer combining of adaptive modulation and coding with truncated ARQ over wireless links, IEEE Trans. Wireless Commun., vol. 3, no. 5, pp. 1746–1755, Sep. 2004.

[52] K. I. Pedersen, G. Monghal, I. Z. Kovács, T. E. Kolding, A. Pokhariyal, F. Frederiksen, and P. Mogensen, "Frequency domain scheduling for OFDMA with limited and noisy channel feedback," in Proc. IEEE VTC, Sep. 2007, pp. 1792–1796.

[53] T. Tang and R. Heath, "Opportunistic feedback for downlink multiuser diversity," IEEE Comm. Letters, vol. 9, pp. 948–950, Oct 2005.

[54] T. Tang, R. W. Heath, Jr., S. Cho, and S. Yun, "Opportunistic feedback in clustered OFDM systems", in proceedings of International Symposium on Wireless personal Multimedia Communications, San Diego, CA, Sep. 2006.

[55] S. Patil and G. de Veciana, "Feedback and opportunistic scheduling in wireless networks", IEEE Trans. Wireless Commun., vol. 6, no. 12, pp. 4227-4238, Dec. 2007.

[56] R. Agarwal, V. Majjigi, Z. Han, R. Vannithamby, and J. Cioffi, "Low complexity resource allocation with opportunistic Feedback over Downlink OFDMA Networks", IEEE JSAC Special Issue on Limited Feedback, Vol. 26, No. 8, Oct. 2008, pp. 1462-1472

[57] S. Jaewoo and J.M Cioffi, "Feedback reduction scheme for downlink multiuser diversity", IEEE Trans. Wireless Commun., vol. 8, no. 2, pp. 668-672, Feb. 2009