



Opportunistic Scheduling with Limited Feedback in Wireless Communications Systems

BY

MOHAMED E. ELTAYEB

A Thesis Presented to the
DEANSHIP OF GRADUATE STUDIES

KING FAHD UNIVERSITY OF PETROLEUM & MINERALS

DHAHRAN, SAUDI ARABIA

In Partial Fulfillment of the
Requirements for the Degree of

MASTER OF SCIENCE

In


TELECOMMUNICATION ENGINEERING

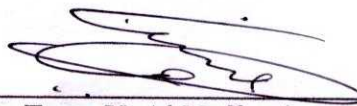
March 2009

DEANSHIP OF GRADUATE STUDIES

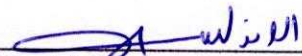
This thesis, written by **MOHAMED E. ELTAYEB** under the supervision of his thesis advisor and approved by his thesis committee, has been presented to and accepted by the Dean of Graduate Studies, in partial fulfillment of the requirements for the degree of **MASTER OF SCIENCE IN TELECOMMUNICATION ENGINEERING**

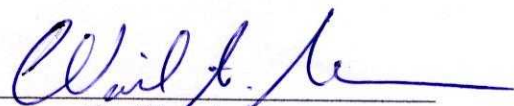
Thesis Committee

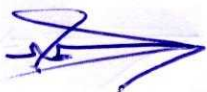

Dr. Yahya Al-Harhi
(Chairman)

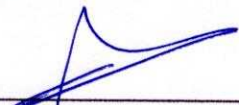

Dr. Tareq Y. Al-Naffouri
(Member)


Dr. Salam A. Zummo
(Member)


Dr. Mohamed A. Landolsi
(Member)


Dr. Wail A. Mousa
(Member)


Dr. Samir H. Abdul-Jauwad
Department Chairman


Dr. Salam A. Zummo
Dean of Graduate Studies

15/4/09
Date



Dedicated to

My Beloved Parents

and

Brothers and Sisters

ACKNOWLEDGEMENTS

First thankfulness and praises are for Allah the Almighty for His countless blessings and ever-enduring mercies.

I would like to express my deepest gratitude to my advisor, Dr. Yahya Al-Harhi, for his continuous guidance, valuable advice and immense efforts. His extensive knowledge and understanding in the field of wireless communications have been invaluable. I also thank him for our fruitful discussions and his availability whenever and wherever I needed him. He treated me more like a friend and a brother. This encouraged me to ask him many questions and to discuss ideas at any time. He always showed great generosity and patience towards me. I thank him for that.

It was a great privilege to work with Dr. Tareq Al-Naffouri. I learned immensely from his theoretical and practical knowledge, and I thank him for his patience and guidance.

I would also like to thank my thesis committee members Dr. Tareq Al-Naffouri, Dr. Salam Zummo, Dr. Wail Mousa and Dr. Mohammad Landalusi for their valuable time devoted to reviewing and correcting my thesis.

Acknowledgment is due to King Fahd University of Petroleum and Minerals for the solid education it gave me and for supporting this research. My stay and experience in the university is unforgettable.

I would like to thank Dr. Ibn Al Waleed Hussein, Dr. Ibrahim El Amin and Dr. Saleh Dofaa for their recommendations to join KFUPM. Also I would like to thank Dr. Abdalla Al-Ahmari and all KFUPM EE faculty and staff members for their cooperation, advice and kind support.

Many thanks and appreciation go to my colleagues for their help and cooperation. I would like to thank Abdullahi Masoud, Motasim Hassan, Mohammed Abdulatif, Umair Mansour, Mujahid Faiz, Ahmed Salim, Hatim Dafallah and all of the 902 and 903 group for their company, encouragement and support.

Finally, I would like to thank my parents, brothers, sisters and relatives for their unreserved support and love that give me the strength to face challenges in my life.

Contents

Acknowledgements	i
List of Tables	vi
List of Figures	vii
Nomenclature	x
Abstract (English)	xv
Abstract (Arabic)	xvi
1 Introduction	1
1.1 Background	2
1.1.1 The Wireless Channel	2
1.1.2 Classical Diversity Techniques	5
1.1.3 Adaptive Modulation	7
1.2 Literature Survey	8

1.3	Thesis Contributions	13
1.4	Thesis Organization	14
2	Opportunistic Scheduling	15
2.1	Introduction	15
2.2	Multiuser Diversity	18
2.3	Scheduling Algorithms	21
2.3.1	Non-Opportunistic Algorithms	22
2.3.2	Opportunistic Algorithms	22
2.4	Conclusion	28
3	Opportunistic Scheduling in Single-Carrier Systems	29
3.1	Introduction	29
3.2	System Model	31
3.3	Scheduling Algorithm	33
3.4	Performance Analysis	36
3.4.1	Feedback Load	36
3.4.2	System Capacity	39
3.4.3	Scheduling Delay	41
3.4.4	System Throughput	42
3.5	Simulation and Numerical Results	42
3.6	Conclusion	54

4	Opportunistic Scheduling in Multi-Carrier Systems	55
4.1	Introduction	55
4.2	System Model	57
4.3	Scheduling Algorithms	60
4.3.1	Algorithm 1	60
4.3.2	Algorithm 2	62
4.4	Performance Analysis	68
4.4.1	Feedback Load	71
4.4.2	Average Spectral Efficiency	81
4.4.3	Probability of Access	84
4.4.4	Scheduling Delay	85
4.4.5	Average System Throughput	86
4.5	Numerical Examples	86
4.6	Conclusion	102
5	Conclusions and Future Research	103
5.1	Thesis Summary and Conclusions	103
5.2	Future Research	105
	Bibliography	106
	Vita	114

List of Tables

3.1	List of selected modulation levels ($\text{BEP}_o = 10^{-4}$)	33
3.2	List of used parameters	42
4.1	List of used parameters	60
4.2	Probing thresholds and the corresponding fed back quantized values for the 3 algorithms approximated over 500 iterations. $K=26$, $L=$ 13 , $\text{BEP}_o = 10^{-4}$ and $\tilde{\gamma} = 15$ dB.	66

List of Figures

2.1	Time varying channel of two users undergoing Rayleigh fading.	17
2.2	Multiuser diversity gain for Rayleigh and Ricean fading channels with Ricean factor $\kappa = 5$ and average SNR = 0 dB.	20
2.3	Average spectral efficiency for MCS, PFA, ORR, RR.	25
2.4	Normalized feedback load for MCS, PFA, ORR, RR.	26
2.5	Time-slot fairness for 4 <i>i.i.d.</i> users with $\bar{\gamma} = 15$ dB.	27
3.1	Structure of a time-slot which consists of a guard period, consisting of a message broadcast slot and k feedback mini slots and a data transmission period.	32
3.2	General flowchart describing the binary feedback algorithm.	37
3.3	Average spectral efficiency for 30 users.	44
3.4	Normalized average feedback load for 30 users.	45
3.5	Average feedback channel payload for 30 users.	46
3.6	Average guard time (msec) for 30 users.	47

3.7	Average feedback channel payload (bits). $\tilde{\gamma} = 20$ dB.	48
3.8	Average guard time (msec). $\tilde{\gamma} = 20$ dB.	50
3.9	System capacity (bits/channel use) for 100 transmitted symbols with $\tilde{\gamma} = 20$ dB.	51
3.10	System capacity (bits/channel use) for 500 transmitted symbols with $\tilde{\gamma} = 20$ dB.	52
3.11	Average system throughput with time-slot durations of 5 and 50 msec. $\tilde{\gamma} = 20$ dB.	53
4.1	System model of the downlink multi-carrier system.	58
4.2	System model with L parallel subchannels and K users.	58
4.3	Flowchart describing Algorithm 1.	63
4.4	Flowchart describing Algorithm 2.	64
4.5	Multi-carrier system time-slot model with 3 subchannels and 3 users.	65
4.6	Markov chain representing algorithm 1. K is the number of users in the system and L is the number of subchannels.	72
4.7	Markov chain representing algorithm 2. K is the number of users in the system and L is the number of subchannels.	82
4.8	Normalized Average feedback load versus the average SNR. $K = 13$	87
4.9	Average spectral efficiency versus the average SNR. $K = 13$	91
4.10	Normalized Average feedback load versus the average SNR. $K = 26$	92

4.11	Average spectral efficiency versus the average SNR. $K = 26$	93
4.12	Average guard time versus the average SNR. $K = 13$	94
4.13	Average guard time versus the average SNR. $K = 26$	95
4.14	Probability of access versus SNR for DSMUDiv scheme, DSMUDiv- EEA scheme and our proposed schemes. $K = 13$	96
4.15	Probability of access versus SNR for DSMUDiv scheme, DSMUDiv- EEA scheme and our proposed schemes. $K = 26$	97
4.16	Average system throughput for a system with $K = 13$, and $t_d = 5$ msec.	98
4.17	Average system throughput for a system with $K = 13$, and $t_d = 50$ msec.	99
4.18	Average system throughput for a system with $K = 26$, and $t_d = 5$ msec.	100
4.19	Average system throughput for a system with $K = 26$, and $t_d = 50$ msec.	101

Nomenclature

Abbreviations

AFL	Average Feedback Load
AM	Adaptive Modulation
AP	Access Point
ASE	Average Spectral Efficiency
ASTH	Average System Throughput
BEP	Bit Error Probability
BPSK	Binary Phase Shift Keying
BS	Base Station
CDF	Cumulative Distribution Function
CNR	Carrier-to-Noise Ratio
CSI	Channel State Information
DSMUDiv	Discrete Rate Switch-Based Multiuser Diversity
DSMUDiv-EEA	DSMUDiv with Enhanced Equal Access
Flash-OFDM	Fast Low-latency Access with Seamless Handoff OFDM

IFFT	Inverse Fast Fourier Transform
ISI	Inter Symbol Interference
JFI	Jain Fairness Index
MASSE	Maximum System Spectral Efficiency
MCS	Maximum CNR Scheduling
M-QAM	Multi-Level Quadrature Amplitude Modulation
OFDM	Orthogonal Frequency Division Multiplexing
ORR	Opportunistic Round-Robin
PDF	Probability Density Function
PFA	Proportional Fair Algorithm
P/S	Parallel-to-Serial
QoS	Quality of Service
RMS	Root Mean Square
RR	Round-Robin
SNR	Signal-to-Noise Ratio
TDMA	Time Division Multiple Access

English Symbols

b_i	Number of bits per constellation
C	System capacity

F	Feedback Load
F_J	Jain Fairness Index
h	Channel gain
K	Total number of users in the system
k	Number of active users
L	Total number of subchannels
l	Subchannel index number
n_k	Zero mean complex Gaussian noise
n	Probing threshold level
N_c	Number of sub-carriers
N	Number of modulation levels
M_k	Modulation level of the k th user
P_{access}	Probability of access
P_n	Selection probability
P_{sub}	Probability of access per subchannel
$q^{(n)}$	Quantized feedback value
q_{kl}	Quantized value fed back by the k th user for the l th subchannel
R	Average spectral efficiency
S	Number of transmitted symbols
S_c	Number of subcarriers per subchannel

t_p	Probing time
T_d	Time slot duration
$u(k, l)$	Scheduling process state with k users and l subchannels
x	Baseband transmitted signal
y_k	Baseband received signal of user k
$w(n, k)$	Scheduling process sub-state with modulation level n and k users
w_k	State modulation level
X_l	Number of users at l_{th} scheduling process

Greek Symbols

σ^2	Time-average power of the received signal before envelope detection
$\Omega(t)$	Envelope of the sum of two quadrature Gaussian noise signals
$\gamma_{th}^{(n)}$	Probing threshold
ρ	Feedback contention probability
τ_g	Guard time
$\tilde{\gamma}$	Average SNR
$\Gamma(k, l)$	Steady-state probability of being in state $u(k, l)$
$\lambda(i)$	Probability that i users are below the minimum threshold
$\Phi_n(k, l)$	Probability of being in substate $w(n, k)$
$\mu_n(k, l)$	Transition probability

α	Substate transition probability parameter
δ	Substate transition probability parameter
β	Substate transition probability parameter
η	Substate transition probability parameter
ν	Substate transition probability parameter
Ψ_n	Steady-state probability

Operators

$E[.]$	Expectation Operator
--------	----------------------

THESIS ABSTRACT

Name: MOHAMED E. ELTAYEB
Title: OPPORTUNISTIC SCHEDULING WITH LIMITED FEEDBACK
IN WIRELESS COMMUNICATIONS SYSTEMS
Degree: MASTER OF SCIENCE
Major Field: TELECOMMUNICATION ENGINEERING
Date of Degree: March 2009

In multiuser systems, multiuser diversity can be exploited by granting the channel resource to the user with the best channel condition. However, to make a scheduling decision, the base station requires the feedback of channel state information of all users. This feedback overhead can increase as the number of users or carriers increases, thus consuming system resources which could be utilized for data transmission. In this thesis, we investigate and propose scheduling techniques that can be employed to reduce this feedback overhead without significant degradation in system performance. To achieve this, we apply quantized feedback information with adaptive modulation and multiple probing thresholds to a polling based system. Closed-form expressions for the average feedback load and the average spectral efficiency are presented. Our numerical results show that our proposed schemes further reduce the feedback load when compared to the optimal scheme under the slow Rayleigh fading assumption. Other parameters such as average guard time, average system capacity, average system throughput, probability of access and scheduling delay are also investigated.

King Fahd University of Petroleum and Minerals, Dhahran.
March 2009

Abstract (Arabic)

ملخص الرسالة

الاسم: محمد الجيلي احمد الطيب

عنوان الرسالة: الجدولة الانتهازية (Opportunistic Scheduling) ذو التغذية الخلفية المحدودة في نظم الاتصالات اللاسلكية

الدرجة: الماجستير في هندسة الاتصالات

التخصص الرئيسي: هندسة الاتصالات

تاريخ التخرج: مارس - 2009م

يمكن استقلال تعدد و تنوع المستخدمين لمنح موارد نظام اتصالات لاسلكي للمستخدم الذي لديه أفضل قناة. لاتخاذ قرار جدولة (Scheduling Decision)، تتطلب القاعدة (Base Station) تغذية خلفية (Feedback) من جميع المستخدمين لمعرفة حالة قنواتهم و بالتالي منح موارد النظام لأفضل مستخدم. تزداد حمولة التغذية الخلفية بزيادة عدد المستخدمين أو الناقلين (Carriers)، و بالتالي زيادة استهلاك موارد النظام التي يمكن استخدامها لنقل بيانات المستخدمين. في هذه الرسالة، نتحرى ونقترح طرق جدولة جديدة لتخفيض حمولة التغذية الخلفية دون تأثير او تدهور ملحوظ في أداء النظام. لتحقيق ذلك، نوظف معلومات التغذية الخلفية الي معلومات محدودة الكم و نطبق التضمين المتكيف وعدده عتبات تحقيق (Probing Thresholds) علي نظام مبني على الاقتراع (Polling Based System).

نتائجنا العديدة توضح ان الخوارزميات المقترحة تقلل من حمولة التغذية الخلفية و متوسط زمن الجدولة عند مقارنتها بالطريقة المثالية و ذلك بعد ما تم افتراض قناة رايلية ذات البط المتلاشي (Slow Rayleigh Fading). تم تقديم صيغ مغلقة لحساب متوسط حمولة التغذية الخلفية و متوسط الكفاءة الطيفية (Average Spectral Efficiency). أيضا تم اعتبار بارامترات أخرى مثل حارس الوقت المتوسط (Average Guard Time)، متوسط سعة النظام (Average System Capacity)، متوسط كفاءة النظام (Average System Throughput)، احتمالية المنح (Access Probability) و زمن تأخير الجدولة (Scheduling Delay).

Chapter 1

Introduction

A communication system basically consists of a transmitter, medium (channel) and a receiver. Such systems, in their simplest forms, have been in existence since the prehistoric era. These systems were characterized by low data rate and high transmission latency. Furthermore, these systems carried small amounts of data over rather short distances. As time passed, the demand for a more reliable and faster communication system increased. This led to the invention of the telegraph and today's communication systems.

In today's typical wireless systems, the base station divides the wireless medium into bands (frequency, time or code). Since the medium is shared, users compete for a system resource. In a time varying channel with a moderate number of users, multi-user diversity can be employed to select the user with the best channel conditions at a time instant [1]. This is achieved by utilizing the natural characteristic of

the fading channel. For a system to always pick the users with the best channel conditions, the channel state information (CSI) of each user is required at the base station. This is known as feedback overhead, and it is a major concern as it can lead to a bottle neck in a high user regime.

1.1 Background

1.1.1 The Wireless Channel

Unlike wired channels (optical fiber, coaxial cable, twisted pair, etc) which are stationary and predictable, wireless channels operate through electromagnetic radiation from the transmitter to the receiver, and they are random and unpredictable. The wireless mobile channel is characterized by the variations of channel coefficients for each user over time and frequency. These variations occur due to the reflection, diffraction and scattering of the electromagnetic waves as they propagate. These effects result in multiple versions of the transmitted signal at the receiver with random phases, amplitudes and arrival times. This effect is known as multipath [2].

The variations in the transmitted signal can be roughly divided into large-scale fading and small-scale fading [3]. Large-scale fading is due to the path loss of the signal as a function of distance and shadowing by large objects such as buildings and hills, and it is typically frequency independent. Small-scale fading is due to the constructive and destructive interference of the multiple signal paths between

the transmitter and receiver, and it is frequency-dependent. Some of the small-scale fading effects due to multipath include [2]

- Rapid changes in the signal strength over a small travel distance or time
- Random frequency modulation due to varying Doppler shifts on different multipath signals
- Time dispersion caused by multipath propagation delays.

Time dispersion due to multipath causes the transmitted signal to undergo either flat or frequency-selective fading. In flat fading, the channel experiences constant channel gain and linear phase response over a bandwidth which is greater than the bandwidth of the transmitted signal. The reciprocal bandwidth of the transmitted signal should also be much larger than the multipath time delay spread of the channel. In frequency-selective fading, the channel possesses a constant gain and linear phase response over a bandwidth that is smaller than the bandwidth of the transmitted signal. The signal also undergoes frequency-selective fading if the symbol time is less than the multipath delay spread. Such fading leads to intersymbol interference at the receiver [2].

Depending on how the transmitted baseband signal changes as compared to the rate of change of the channel, the channel may be classified as fast fading or slow fading. In a fast fading channel, the channel impulse response changes rapidly within the symbol duration. That is, the coherence time of the channel is smaller than the

symbol duration. This leads to signal distortion at the receiver. In slow fading, the channel impulse response changes at a rate slower than the transmitted signal, i.e. the symbol time is smaller than the coherence time.

Rayleigh and Ricean Fading Distributions

Rayleigh fading distribution is usually used to describe the statistical time-varying nature of the received signal envelope. When the composite received signal consists of a large number of plane waves (multi-path), the central limit theorem can be applied, and the received complex envelope $g(t) = g_I(t) + jg_Q(t)$ can be treated as a complex Gaussian process. Under these conditions, the envelope of the sum of two quadrature Gaussian noise signals, $\alpha(t) = |g(t)|$, obeys a Rayleigh distribution. The PDF of the Rayleigh distribution is given by

$$p_\alpha(r) = \frac{r}{\sigma^2} e^{-\frac{r^2}{2\sigma^2}}, \quad 0 \leq r \leq \infty \quad (1.1)$$

where σ is the RMS value of the received voltage signal before envelope detection, and σ^2 is the time-average power of the received signal before envelope detection. The average envelope power $E[\alpha^2] = \Omega_p = 2\sigma^2$ and the corresponding PDF is

$$p_\alpha(r) = \frac{2r}{\Omega_p} e^{-\frac{r^2}{\Omega_p}}, \quad 0 \leq r \leq \infty. \quad (1.2)$$

The corresponding squared envelope is exponentially distributed with the following density [3] [2] [10]

$$p_{\alpha^2}(r) = \frac{1}{\Omega_p} e^{-\frac{r}{\Omega_p}}, \quad 0 \leq r \leq \infty. \quad (1.3)$$

When there is a dominant line-of-sight component with the other scattered and reflected components, the fading envelope distribution is known to be Ricean. As the dominant signal becomes weaker, the composite signal resembles a noise signal which has a Rayleigh envelope. The pdf of the Ricean distribution is given by [2]

$$p(r) = \frac{r}{\sigma^2} e^{-\frac{r^2+A}{2\sigma^2}} I_0\left(\frac{Ar}{\sigma^2}\right), \quad \text{for}(A \geq 0, r \geq 0) \quad (1.4)$$

where A is a parameter that denotes the peak amplitude of the dominant signal and $I_0(\cdot)$ is the modified Bessel function of the first kind and second order. The Ricean distribution is often described in terms of a parameter κ which is defined as the ratio between the deterministic signal power and the multipath variance. κ is given by $\kappa = A^2/(2\sigma^2)$.

1.1.2 Classical Diversity Techniques

Reliable communication depends on the strength of a signal as it propagates. Diversity is a powerful technique that provides link improvement at low cost. In fading channels, there is high probability that a path will be in a deep fade at a time instant, and thus the path will suffer from errors. A natural way to combat these

errors is to reduce the probability of a signal fade at an instant of time. This is done by providing multiple versions of the transmitted signal at the receiver, thus ensuring reliable communication as long as one of the received versions is not in a deep fade. This technique is called diversity, and it can dramatically improve the performance over fading channels. Below are some of the most common diversity techniques.

Frequency Diversity

In this technique, the users' data is transmitted in more than one frequency. The difference between the frequencies must be more than the coherence bandwidth, so that they can be independent.

Time Diversity

In time diversity, another replica of the signal is sent after a certain time interval, which is more than the coherence time. This time separation is required in order to ensure that the channel characteristics have changed enough and the transmitted signals are uncorrelated.

Antenna Diversity

Antenna diversity is obtained by placing multiple antennas at the transmitter and/or the receiver. In either case, the separation distance between the antennas must be

more than the coherence distance which depends on the scattering environment and on the carrier frequency [2]. In this case, the channel gains between the different antenna pairs fade more or less independently, and independent signal paths are created. The receiver or transmitter can then choose the strongest path for data transmission.

1.1.3 Adaptive Modulation

In the wireless channel, the user's signal-to-noise ratio (SNR) depends on a number of factors including the distance between the users and the base stations, path loss exponent, shadowing, short-term Rayleigh fading and noise. In order to improve system capacity, peak data rate and coverage reliability, the signal transmitted to and by a particular user is modified to account for the signal quality variation through a process commonly referred to as link adaptation. Adaptive modulation offers a link adaptation method that promises to raise the overall system capacity. Adaptive modulation provides the flexibility to match the modulation scheme to the average channel conditions for each user, i.e. provide a large constellation when channel quality is high and a small constellation when the channel quality is low. This is done by changing the modulation format to suit the current SNR of the user. The implementation of adaptive modulation offers several challenges. In order to select the appropriate modulation, the scheduler must be aware of the channel quality. Errors in the channel estimate will cause the scheduler to select the wrong

modulation scheme and either transmit at a high power, wasting system resources, or at a low power, raising the error rate.

1.2 Literature Survey

Opportunistic communication maximizes the spectral efficiency by measuring when and where the channel is good and it transmits only in those degrees of freedom [3]. An important scheduling criterion is the feedback load, as it contributes to the overall system throughput degradation. The main objective is to keep feedback load to a minimum and to maintain a satisfying Quality of Service (QoS).

A thorough discussion on feedback reduction was conducted in [4] (and references therein). For instance, in [5], users compared their instantaneous SNR with a predetermined threshold, and only users who had channel qualities above the threshold were allowed to feedback while the others remained silent. The threshold was optimized to meet a specified outage probability. In case no user fed back, a random user was chosen and given the channel resource. This technique reduced the feedback load, as only a subset of users were allowed to feedback. However, the random selection when an outage occurs can result in some capacity loss. The work in [5] was extended in [6] where full feedback was required when an outage occurred instead of just selecting a random user. This improved the system capacity as compared to [5] but it increased the feedback load. Similar work, which used

multiple thresholds to reduce the feedback load, was proposed in [7]. Only users that were above a certain threshold were allowed to feedback in a contention-based feedback channel. This reduced the feedback load at the expense of scheduling delay. The effect of feedback quantization on the throughput of multiuser diversity was studied in [14] where it was concluded that only a few quantization levels were required to capture most of the diversity gain. The work in [15] showed that one-bit fixed rate feedback was able to achieve the optimal capacity growth rate. However, there was a slight loss in capacity due to the low resolution feedback. Similarly, [16] considered one-bit feedback with imperfect feedback channel, where all users above a particular threshold were allowed to feedback one-bit information. Users were then divided into two sets, and the channel resource was allocated to one user belonging to the set which reported favorable channel conditions. This scheme also reduced the feedback rate and load, with slight loss in capacity due to low resolution of one-bit quantization. Feedback quantization was also studied in [8] where the authors considered a more practical model which implemented discrete rates. The authors considered a probing system in which users were probed with a defined threshold and were allowed to feedback their quantized channel state information (CSI). The first user that reported a channel state that was equal to or higher than the threshold was given the channel resource. When compared with the optimal selective diversity scheme, their scheme reduced the feedback load with no loss in spectral efficiency. Although this scheme reduced the feedback load at a mid to high

average SNR regime, full probing (full feedback) was required at a low average SNR. The authors also extended their research to multi-carrier systems [17] where each scheduled user was not allowed to compete for another subchannel till all remaining subchannels were scheduled. More feedback reduction was obtained, with a slight loss in capacity due to multiuser diversity gain loss. An ALOHA-based multi-carrier system was adopted in [18] where each user compared all of its OFDM sub-carriers with an optimized threshold and fed back only for those that exceeded the threshold. An opportunistic scheduling scheme with partial channel state information was considered in [19] and [21]. The scheme required partial feedback instead of full feedback, by allowing users that were above a threshold to feedback channel state information for a number of predetermined channels or a group consisting of the best channels only. This reduced the feedback load with slight capacity loss. Similarly, [20] introduced a scheme in which users compared all of their sub-carriers with a capacity threshold. If the sum-capacity of the user was above the threshold, the user was given the channel resource. Otherwise, the next user was examined till the last user. The last user, or the user with the best sum-capacity, was given the channel resource. This scheme reduced the feedback load at high average SNR, but it required full feedback at low average SNR. Moreover, there was a slight capacity loss when the number of sub-bands increased.

In addition, the work in [23]-[28] considered a contention-based feedback channel in single-carrier systems to reduce the feedback load. For instance, in [23] and [28],

splitting algorithms with threshold optimization was considered. These algorithms tried to find the best user by splitting users into groups depending on an optimized threshold. In [27], a spread-spectrum contention-based feedback channel was used, where a unique spreading code per user was employed for user identification. It was shown that, as the spreading code increased, the throughput approached that of the full feedback scheme. Another scheme, which required the users to feedback their user identification information only, was proposed in [26]. This scheme used multiple thresholds to find the best user. However, a concern arose when the number of users increased as more feedback was required in order to identify each user. The work in [24] and [25] used a ranked list which was distributed to all users. When users competed for a channel resource, the first user in the ranked list fed back. If the other users in the list did not sense any transmission, they fed back according to their list, until one user above a predetermined threshold fed back.

Other proposed schemes considered the transmission of feedback information for a group of carriers (clusters) only [30] [31] [29]. In particular, each user sent feedback information for the strongest clusters only, instead of all. This reduced the feedback overhead without significant performance loss. This scheme was optimized in [31] where the channel gain threshold and the number of clusters per user were taken into account to improve the system throughput. Furthermore, in [29], only the indices of the S strongest clusters were fed back, instead of feeding back an SNR value for each cluster. Resource allocation was then given to a random eligible user. This random

selection eventually led to performance loss in spectral efficiency [29]. In addition to indexed feedback, adaptive cluster request based on an outage probability was also considered instead of static clustering. This improved the spectral efficiency with the expense of slight outage penalty. Similarly, a QoS-aware selective feedback model was taken into account in [32]. This was done by allowing each user to feedback information for the channel sets that it required by using a target bit error rate (BER). The base station then optimally assigned the channels to users with the objective of maximizing the number of users or the sum of the users' utility values (system capacity). This scheme reduced the feedback overhead. However, it required complex search algorithms that exponentially increased as the shared channel sets increased.

Other related work considered the use of a delta modulation-based scheme in orthogonal frequency-division multiple access systems (OFDMA) [33]. Instead of feeding back accurate channel information for each carrier, each user fed back a code which indicated whether the channel gain was above or below the channel gain of the previous carrier. This scheme reduced the feedback load, with some performance loss due to imprecise channel estimation. The work in [34] dealt with proportional fair scheduling in an OFDMA system to reduce feedback. In the proposed scheme, the scheduler first chose the best k users and then requested each of the k users to feedback their best l subchannels. Subchannel grouping was proposed in [35] where users requested feedback from one or more groups if all of the group's subchannel

gains exceeded a threshold. For feedback, users represented each group with one bit and fed back a compressed vector indicating the desired groups. Similarly, users were assigned to groups, and they requested group allocation by sending identification bits through a contention channel. In spite of feedback reduction, the segregation imposed in both schemes led to diversity gain loss, and the sum capacity depended on the number of subchannel or user groups. Finally, in [36], feedback compression was considered by exploiting the correlation in time and frequency between neighboring subchannels. A lossy compression algorithm was employed by each user prior to feedback transmission. Of course, due to lossy compression of feedback information, there was some loss in throughput.

1.3 Thesis Contributions

The main contributions of this thesis can be summarized as follows:

- We introduce a feedback load and rate reduction algorithm in a single-carrier system, and we derive closed form expressions for the average spectral efficiency and feedback load.
- We present two feedback load reduction algorithms in a multi-carrier system, and we derive closed form expressions for the average spectral efficiency and feedback load for both algorithms.

1.4 Thesis Organization

This thesis is organized as follows. In Chapter 2, we give an overview on some opportunistic scheduling algorithms and we compare their effectiveness in terms of capacity, feedback load and fairness.

In Chapter 3, we consider a feedback reduction scheme in a discrete-rate single-carrier system that reduces the feedback load with no loss in spectral efficiency when compared with the optimal full feedback scheme. We also derive closed-form expressions for the average spectral efficiency and the average feedback load. Furthermore, we investigate other parameters such as average guard time, average system capacity and the average system throughput.

Opportunistic scheduling in a discrete-rate multi-carrier system is considered in Chapter 4. We introduce an adaptive probing threshold algorithm that reduces the feedback load, and then we extend the algorithm by employing a rather relaxed probing mechanism. Closed-form expressions for the average feedback load and average spectral efficiency are then derived for both schemes. Moreover, we study the effect of this feedback load reduction on the system's average spectral efficiency, probability of access, throughput and scheduling delay.

Finally, in Chapter 5, conclusions are drawn and possible future research directions are pointed out.

Chapter 2

Opportunistic Scheduling

2.1 Introduction

Scheduling is a method that permits multiple users to share a common resource (transmit power, bandwidth, modulation scheme, etc.) The scheduler needs to identify all the users and their QoS demands, and then to employ a scheduling algorithm that allocates the system resources efficiently to the users. To support QoS in a packet switching network, a scheduling algorithm should seek these goals [9]

- Sharing bandwidth
- Providing fairness
- Meeting bandwidth guarantees

- Meeting loss guarantees
- Meeting delay guarantees
- Reducing delay variations

These points necessitate specially tailored scheduling algorithms that efficiently utilize the system resources. Hence, the scheduler should opportunistically exploit the users' time varying channels to achieve a higher network capacity. A scheduling algorithm is said to be opportunistic when it takes the channel quality into consideration before a scheduling decision. This means that the scheduler can select the best user, according to how the algorithm solves the trade-off between capacity and QoS/fairness. Figure 2.1 shows the time-varying channels for two users. In this case, if opportunistic scheduling is employed, the overall system capacity can be increased by always scheduling the best user at a time interval (riding the peak.)

The main objective of opportunistic schedulers is to increase the maximum system spectral efficiency (MASSE) which is defined as the maximum average sum of spectral efficiencies within a cell which is shared between all the users [Bits/Sec/Hz]. Two important scheduling criteria are feedback overhead and fairness. Feedback overhead is the result of the channel state information of all users required by opportunistic schedulers. Moreover, if the scheduler always serves the user with the best channel conditions, users with poor channel conditions may lag behind with no service. Unlike fair scheduling algorithms, opportunistic schedulers are classified as

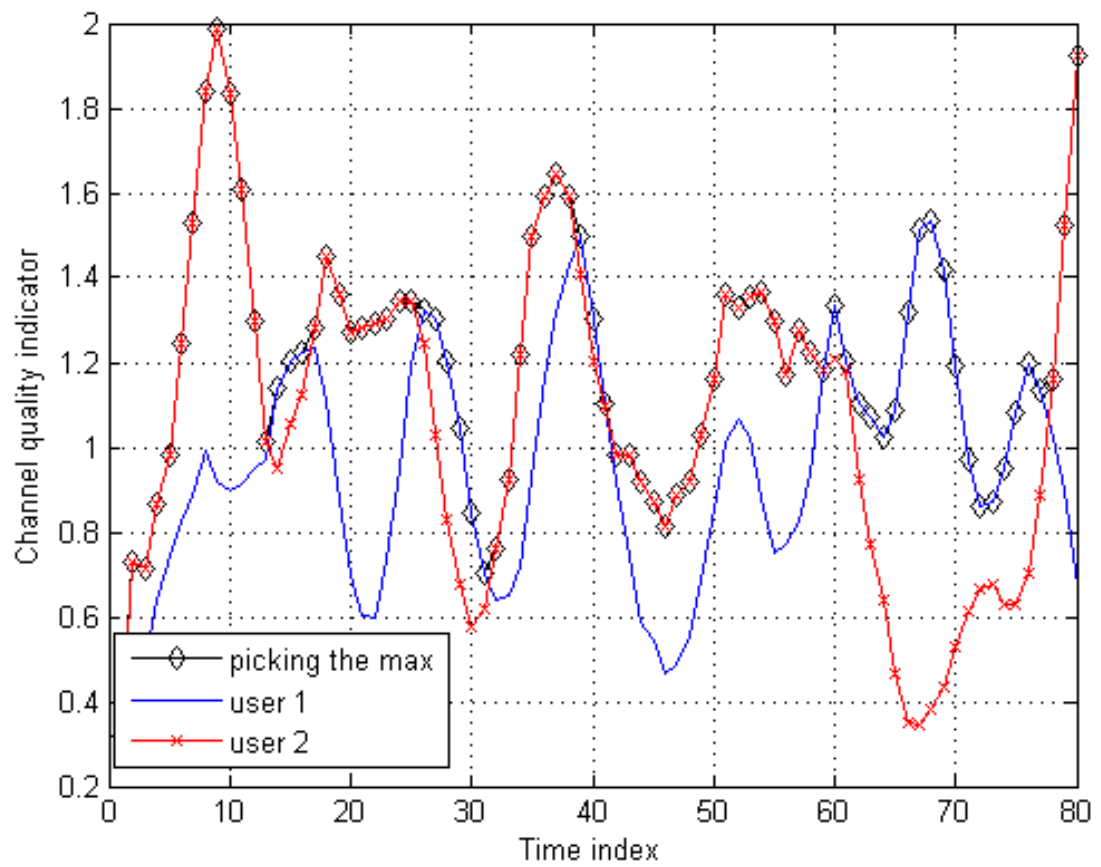


Figure 2.1: Time varying channel of two users undergoing Rayleigh fading.

greedy schedulers, unless a fairness measure is included, as they always schedule the user with the best channel condition.

2.2 Multiuser Diversity

Channel fading was traditionally viewed as a source of error and unreliability that is undesirable, but fading is now employed as a requirement to achieve a multiuser diversity gain.

Multi-user diversity is a way to exploit channel conditions by selecting one from an array of connected users [1]. The users (mobile or stationary) may have different noise, path loss, shadowing and multipath fading, and their instantaneous signal-to-noise ratio (SNR) will reach peak values at different times independently if the channels are uncorrelated. Although diversity in time, space and frequency provides a large diversity gain for single users, multi-user diversity can be used to maximize the overall average throughput of the system. As the number of users that fade independently increases, there is a high probability of finding a user with good channel conditions at a time instant, thus increasing the multi-user diversity gain which results in better utilization of the system resources.

The amount of multiuser diversity depends on the tail of the fading distribution $|\alpha|^2$ [3]. The heavier the tail, the more probably there is a user with a very strong channel, and the larger the multiuser diversity gain. Figure 2.2 compares the

capacity for users undergoing Rayleigh and Ricean fading and users with AWGN only. As seen, AWGN channel has the minimum capacity as the channel is almost constant and does not experience much randomness. Ricean channel has a dominant path and thus it is less random and a lighter tail when compared to the Rayleigh distribution. As a consequence, it has a smaller diversity gain when compared to the Rayleigh case.

To achieve multiuser diversity gains, some systems aspects should be taken into consideration. For instance, the base station requires channel quality measurements of all users, and it should be able to schedule users according to their channel qualities. Fortunately, these features are already available in the designs of many third-generation systems. Other issues that need to be addressed include:

- **Fairness and Delay:** The idea of always selecting users that have good channel conditions can deprive weak users that are either far from the base station or do not have enough scatters in their environments. Additionally, there might be some latency prior to the scheduling decision.
- **Channel measurements and Feedback:** The base station should have accurate error-free CSI from all users to exploit multiuser diversity gain. Another concern arises when the number of users increases, as the base station should be able to handle the large amount of feedback traffic which may lead to a bottleneck.

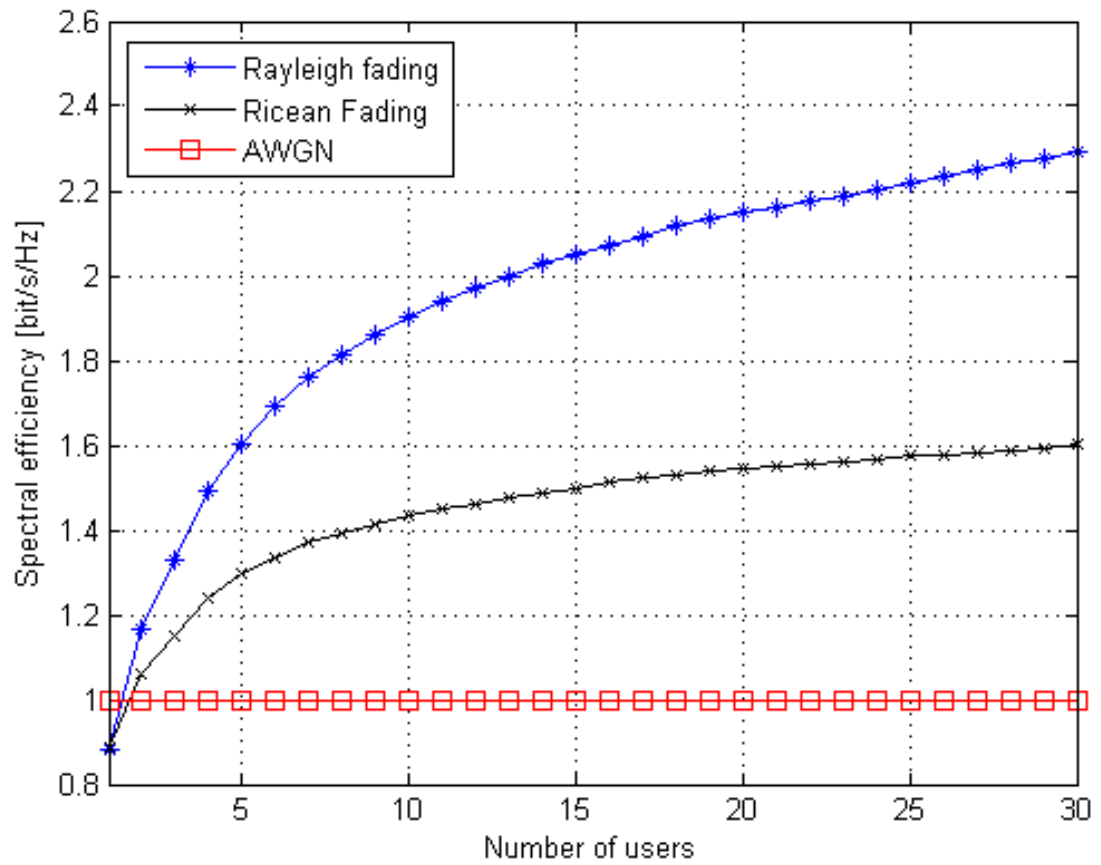


Figure 2.2: Multiuser diversity gain for Rayleigh and Ricean fading channels with Ricean factor $\kappa = 5$ and average SNR = 0 dB.

- Slow and limited fluctuations: As observed, multiuser diversity gains depend on the distribution of the channel fluctuations. In particular, larger and faster variations are preferred over slow ones. However, the channel may fade very slowly compared to the delay constraint of the application, and so transmission cannot wait until the channel reaches its peak.

A key challenge is to address these issues while at the same time exploiting multiuser diversity gain.

2.3 Scheduling Algorithms

In this section, we briefly give an overview on some of the scheduling algorithms. Basically, we consider non-opportunistic algorithms, opportunistic algorithms with no fairness constraints, and opportunistic algorithms with a fairness constraint. For fairness comparison, we employ the *Jain Fairness Index* (JFI) [37][38] which was used in several recent papers. JFI ($F_J(T)$) is given by

$$F_J(T) = \frac{(E_T[X])^2}{E_T[X^2]} \quad (2.1)$$

where X is a random variable describing the amount of resource allocated to a user and $E_T[.]$ is the expectation calculated over T time-slots.

2.3.1 Non-Opportunistic Algorithms

As stated previously, algorithms that do not take channel conditions into considerations are not considered opportunistic. An example of this algorithm is the traditional Round-Robin (RR) algorithm used in conventional time-division multiple-access (TDMA) systems. To schedule a user, the scheduler simply schedules the first user in the queue followed by the second and so on, i.e. the scheduled user's index is

$$i^* = \begin{cases} i(t-1) + 1 & , i(t-1) = 1, 2, \dots, K-1 \\ 1 & , i(t-1) = K. \end{cases} \quad (2.2)$$

As seen in Figures 2.3, 2.4 and 2.5, the RR algorithm does not contribute to any feedback overhead, and it has the best fairness index (similar to ORR). These merits are achieved at the expense of low spectral efficiency as compared to opportunistic schemes.

2.3.2 Opportunistic Algorithms

Opportunistic algorithms take channel conditions into consideration before the resource allocation to any user. As a result, they yield better system performance at the expense of feedback load overhead and fairness if a fairness constraint is not imposed. This is clearly shown in Figures 2.3, 2.4 and 2.5.

Greedy Algorithms

Greedy algorithms are rate-optimal algorithms that do not take fairness into account. These algorithms always schedule the user with the best channel conditions. An example of these algorithms is the *Maximum CNR Scheduling* (MCS) algorithm [12]. In the MCS algorithm, the scheduler requests all users to report their carrier-to-noise ratios (CNR) to the base station. Having done so, the scheduler then schedules the user with the highest possible data rate. The scheduled user index can be expressed as

$$i^* = \arg \max_i C_i(t) \quad (2.3)$$

where $C_i(t)$ is the capacity.

Opportunistic Algorithms with Fairness Constraints

As seen in the previous section, greedy algorithms do not consider fairness constraints. We consider two opportunistic algorithms that provide some degree of fairness.

Opportunistic Round Robin (ORR)

This algorithm is similar to the RR algorithm. Instead of just blindly scheduling users, this algorithm requests channel-state information from the unscheduled users only. This adds a certain amount of feedback overhead which is less than the feedback overhead produced by the MCS algorithm. Fairness is induced here by

removing the user from the list after being scheduled, i.e. the scheduled user cannot request a channel resource unless all other users have been scheduled. The feedback load (F) is given by [38]

$$F = \sum_{k=2}^K k. \quad (2.4)$$

Proportional Fair Algorithm (PFA)

This algorithm tries to maximize the system spectral efficiency with a fairness constraint that does not starve weak users. The scheduler selects a user according to [39]:

$$i^* = \arg \max_i \left(\frac{\gamma_i(t)}{T_i(t)} \right) \quad (2.5)$$

where i is the scheduled user, $\gamma_i(t)$ is the CNR and $T_i(t)$ is the average throughput for user i during the time window. The main objective is to schedule the user with the highest CNR and the lowest average throughput during the time window. This constraint will eventually lead to an equal throughput for all users.

In our proposed algorithms, we employ greedy scheduling basically for two reasons: (i) Our work focuses on feedback overhead reduction with no or minimum efficiency loss. Greedy algorithms provide excellent performance and thus act as a bound in which our proposed algorithms should not deviate much. (ii) The feedback overhead imposed by greedy algorithms acts as an upper bound for all other

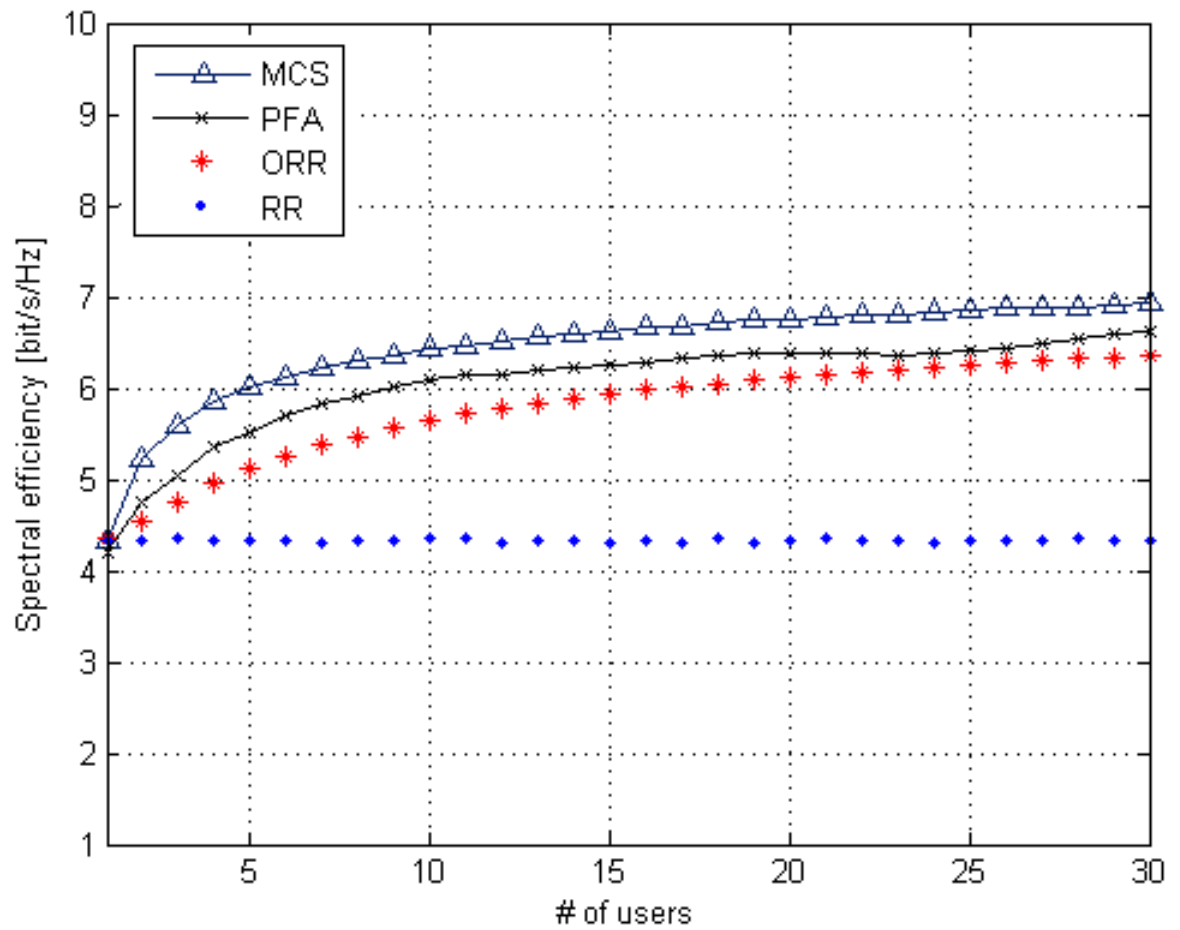


Figure 2.3: Average spectral efficiency for MCS, PFA, ORR, RR.

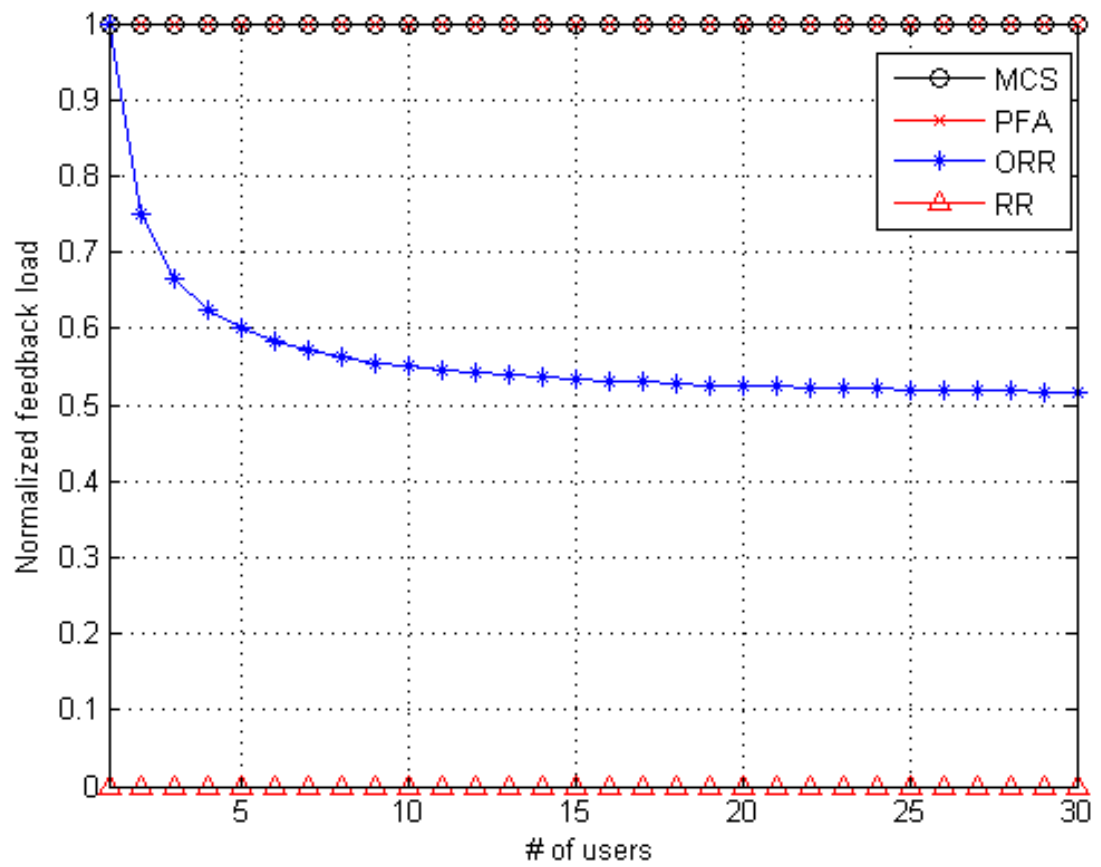


Figure 2.4: Normalized feedback load for MCS, PFA, ORR, RR.

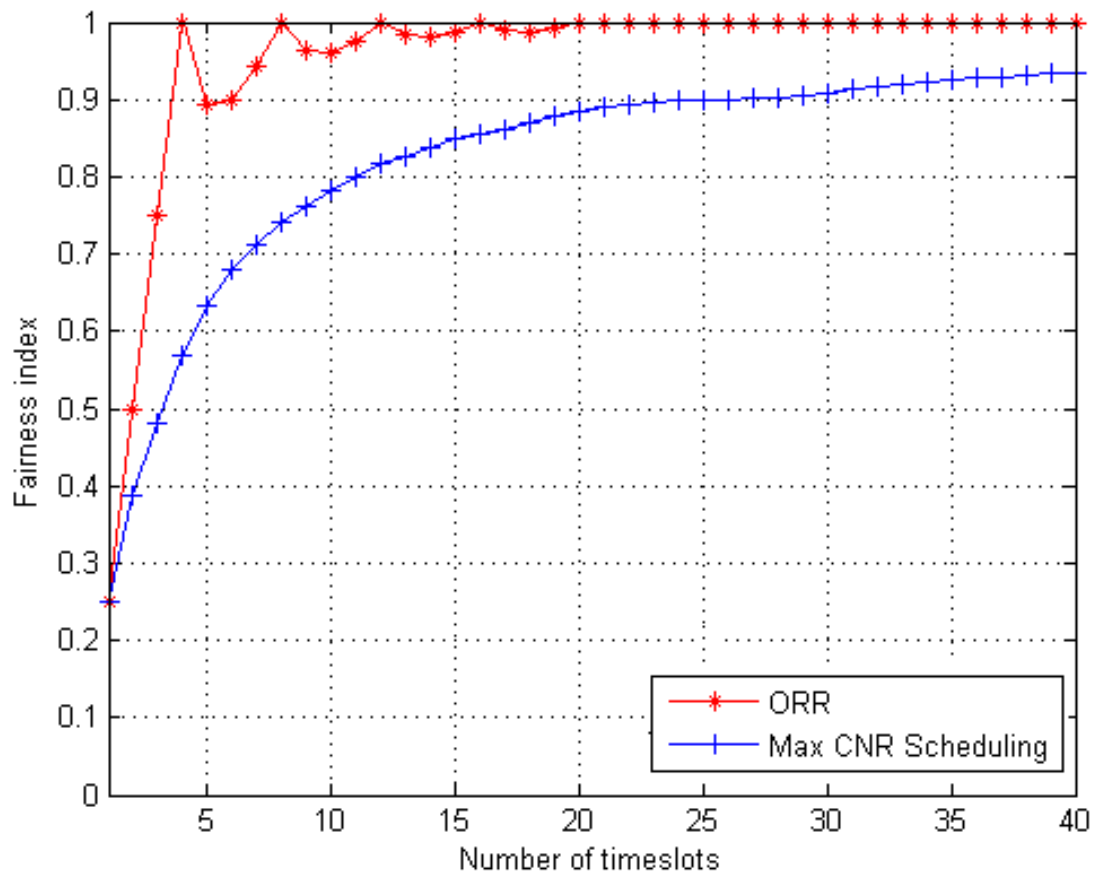


Figure 2.5: Time-slot fairness for 4 *i.i.d.* users with $\bar{\gamma} = 15$ dB.

fair or less greedy algorithms. Any feedback reduction scheme imposed on greedy algorithms can be implemented in other opportunistic schemes.

2.4 Conclusion

In this chapter, we gave a brief insight on opportunistic scheduling algorithms, and we compared their performance and feedback overhead. We also demonstrated the fairness/performance and the feedback overhead/performance trade-off. We showed that greedy algorithms utilize the system capacity efficiently by always scheduling the best user. However, this could lead to QoS violation in some networks, as some users may experience large delays before being scheduled. Furthermore, greedy algorithms are dominated by strong users. This could starve users with low SNR and thus lead to QoS violation in some networks. To counter the effect of greedy algorithms, we showed that a fairness constraint could be added to greedy algorithms to give opportunity to users with weak SNR. This comes at the expense of slight performance loss.

Chapter 3

Opportunistic Scheduling in Single-Carrier Systems

3.1 Introduction

The rapid demand for wireless broadband communication services calls for efficient algorithms that satisfy QoS requirements with minimum service delay. Traditionally, fading was considered as a serious channel impairment that must be mitigated, but now it is considered as a major requirement and sometimes induced to increase the system capacity [1]. In the presence of fading, with a reasonably large number of users, the scheduler can exploit multiuser diversity by requesting feedback information from all users. Clearly, this technique is expensive in terms of spectrum resource, and it may be impractical when the number of simultaneously active

users becomes high due to the large amount of feedback information [13].

In this chapter, we propose a scheduling scheme, similar to the discrete rate switch-based multiuser diversity (DSMUDiv) scheme [8], that reduces the feedback load with no penalty loss in spectral efficiency when compared to the optimal selective diversity scheme (full feedback). The DSMUDiv scheduling scheme reduces the feedback load by using quantized feedback and adaptive modulation (AM). One of its drawbacks is that it requires full feedback load at low average SNR. In our proposed scheme, we extend the work by applying multiple thresholds and a binary probing mechanism that requires only a binary feedback channel. The following highlights the advantages of our scheme :

- Reduced feedback rate due to one-bit feedback.
- Reduced feedback load with no loss in average spectral efficiency when compared to the optimal scheme.
- Improved system throughput due to reduced scheduling delay.
- Improved system capacity due to reduced feedback load and rate.

Our simulation and numerical results show that our scheme further reduces the feedback load when compared to both the DSMUDiv scheme and the optimal scheme under slow Rayleigh fading assumption. We also investigate other parameters such as the guard time, system capacity and system throughput.

The remainder of this chapter is organized as follows. Section 3.2 introduces the system model. In sections 3.3 and 3.4 we present the scheduling algorithm and the performance analysis of the algorithm respectively. We show numerical and simulations results in section 3.5, and finally we explain our conclusions in section 3.6.

3.2 System Model

The system, with a single antenna at the transmitter, serves K users by a single base station (BS). The channel is assumed to be experiencing flat Rayleigh fading, and the users are independent and identically distributed (*i.i.d.*). The baseband channel model is

$$y_k(T) = h_k(T) \cdot x(T) + n_k(T); \quad k = 2, 3, \dots, K, \quad (3.1)$$

where $x(T) \in C$ is the transmitted signal in time-slot T and $y_k(T) \in C$ is the received signal of user k in time-slot t . $x(t)$ is assumed to have the same constant normalized transmitted power over time, i.e. $E(|x(t)|^2) = 1$. The noise processes $n_k(t)$ are *i.i.d.* sequences of zero mean complex Gaussian noise with variance σ^2 . The fading channel gain from the base station to the k_{th} user in time-slot t is $h_k(t)$. Here, scheduling at the downlink channel is considered.

Scheduling is organized on a time-slot basis. As shown in Figure 3.1, a time-slot consists of a guard period, which contains a message broadcast slot and several

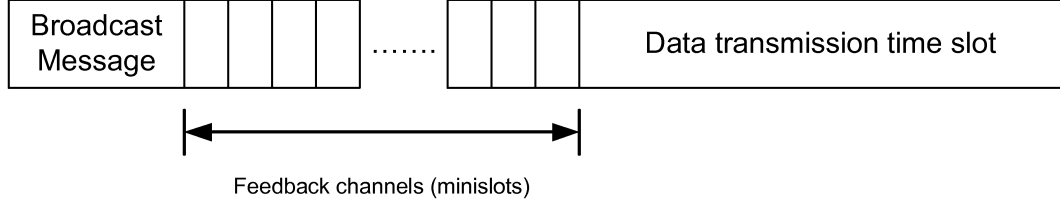


Figure 3.1: Structure of a time-slot which consists of a guard period, consisting of a message broadcast slot and k feedback mini slots and a data transmission period.

mini slots used for feedback, and a transmission period. This time-slot is assumed to be shorter than the coherence time of the channel. The probing of the users takes place during the guard time duration, which is between the bursts. The probed users estimate their downlink channel's SNR, and they feedback a binary acknowledgement indicating whether or not their channels are above the probing threshold. It is assumed that the SNR estimation is perfect, that the system can perfectly detect collisions, and that the feedback channel is error free. Multi-level quadrature amplitude modulation (M-QAM) [41] is used to modulate the user's data depending on the user's channel quality. The probing thresholds, according to [8] (see Table 3.1), are given by:

$$\begin{aligned}\gamma_{th}^{(1)} &= [\text{erfc}^{-1}(2 \cdot \text{BEP}_o)]^2, \\ \gamma_{th}^{(n)} &= -\frac{2}{3}(2^n - 1) \ln(5 \cdot \text{BEP}_o); \quad n = 2, 3, \dots, N \\ \gamma_{th}^{(N+1)} &= +\infty,\end{aligned}$$

Table 3.1: List of selected modulation levels ($\text{BEP}_o = 10^{-4}$)

Modulation Level	Switching Threshold (dB)
BPSK	$\gamma_{th}^{(1)} = 6.9$
4-QAM	$\gamma_{th}^{(2)} = 11.8$
16-QAM	$\gamma_{th}^{(3)} = 18.8$
64-QAM	$\gamma_{th}^{(4)} = 25$

where BEP_o is the average target bit error probability, and $\text{erfc}^{-1}(\cdot)$ denotes the inverse error function. Assuming discrete rates, similar to [8], the number of threshold values depends on the number of the modulation levels M_n (where $M_n < M_{n+1}$ and $0 \leq n \leq N$) used at the transmitter. M_0 is equal to 1, and it indicates that the user is in a deep fade. M_N is the highest modulation level, and N is the total number of threshold or modulation levels. The choice of modulation is taken according to the threshold level $\gamma_{th}^{(n)}$ for which the acknowledgement was received. For instance, if γ_k is the estimated SNR of the k th user, then the modulation level of the k th user (M_k) is

$$M_k = \begin{cases} M_0 & \text{if } \gamma_k < \gamma_{th}^{(1)} \quad (\text{outage}) \\ M_n & \text{if } \gamma_{th}^{(n)} \leq \gamma_k < \gamma_{th}^{(n+1)} \\ M_N & \text{if } \gamma_k \geq \gamma_{th}^{(N)} \end{cases} \quad (3.2)$$

3.3 Scheduling Algorithm

In this section, we give a detailed description of the proposed algorithm. The algorithm guarantees that the best user is always selected and has a minimum feedback

load of 2 and a maximum of $\frac{K}{2} + N$. The algorithm achieves this by using a two-stage scheduling process. The first stage gives the scheduler an estimate on which threshold level the best user is likely to be found, and the second stage is used to find the best user.

Generally, in the first stage (query mode), all users are allowed to transmit a 1 if they are above or at a predetermined broadcasted threshold. Once one or more users feedback, the scheduler then goes into the second stage (search mode). If no user responds, the threshold is sequentially lowered. Unlike [8], the base station starts probing two random users at a time, instead of one in the search mode. Here, we assume all users have a unique identity number (ID) and the probing request is heard by all users. When the two randomly selected users are probed, one of the following scenarios may occur: (i) Only one user has an instantaneous SNR above or equal to the threshold. In this case, that user will feedback a 1 if it has a higher rank than the other user, or a -1 if it has a lower rank. The channel is granted to that user, and the guard period is terminated. (ii) Both users have an instantaneous SNR above or equal to the threshold. In this case, both users feedback and a collision occurs. As a result, the scheduler simply selects one of the two users randomly and then terminates the guard period. (iii) No users have an instantaneous SNR above or equal to the threshold. In this situation, none of the users will respond. The scheduler waits for one mini-slot duration and then allows another pair of users to contend in another mini-slot.

A general description of the algorithm is illustrated in the flowchart given in Figure 3.2. Below, we give a detailed description of our algorithm.

- The scheduler (base station) sequentially arranges all the users and assigns each user a unique ID.
- The base station broadcasts a query message (first stage) to all users with the highest threshold level, γ_{th}^N , allowing all users that are above or equal to the threshold to feedback a 1 in one mini-slot with probability $\rho = 1$.
- If no user feeds back, the threshold is sequentially lowered until at least one user feeds back or the lowest threshold level is reached.
- If one or more users feedback a 1, then the scheduler knows that at least one user has an instantaneous SNR lying within the broadcasted threshold level, and thus goes into the second stage (search mode).
- The scheduler then randomly probes two users and allows them to contend for a new mini-slot with probability $\rho = 1$.
- Assuming that the probing request is heard by all users, the user with the higher index number feeds back a 1 if it has an instantaneous SNR above or equal to the threshold. If the SNR is below the threshold, the user remains silent.

- The user with the lower index feeds back a -1 if it has an instantaneous SNR above or equal to the threshold, but it will remain silent if the SNR is below the threshold.
- If none of the users are above the threshold, both users will remain silent during the mini-slot duration.
- If one of the users feeds back successfully, then the user is identified and given the channel resource.
- If a collision occurs, then the channel resource is given randomly to either of the two users.
- If neither of the two users responds, then another set of two users are allowed to contend for another mini-slot.

3.4 Performance Analysis

In this section, we evaluate the performance of the scheduling algorithm in terms of feedback load, system capacity, scheduling delay and overall system throughput.

3.4.1 Feedback Load

The average feedback load (AFL) is defined as the average number of consumed mini-slots until a user is scheduled. The feedback load in this case consists of two

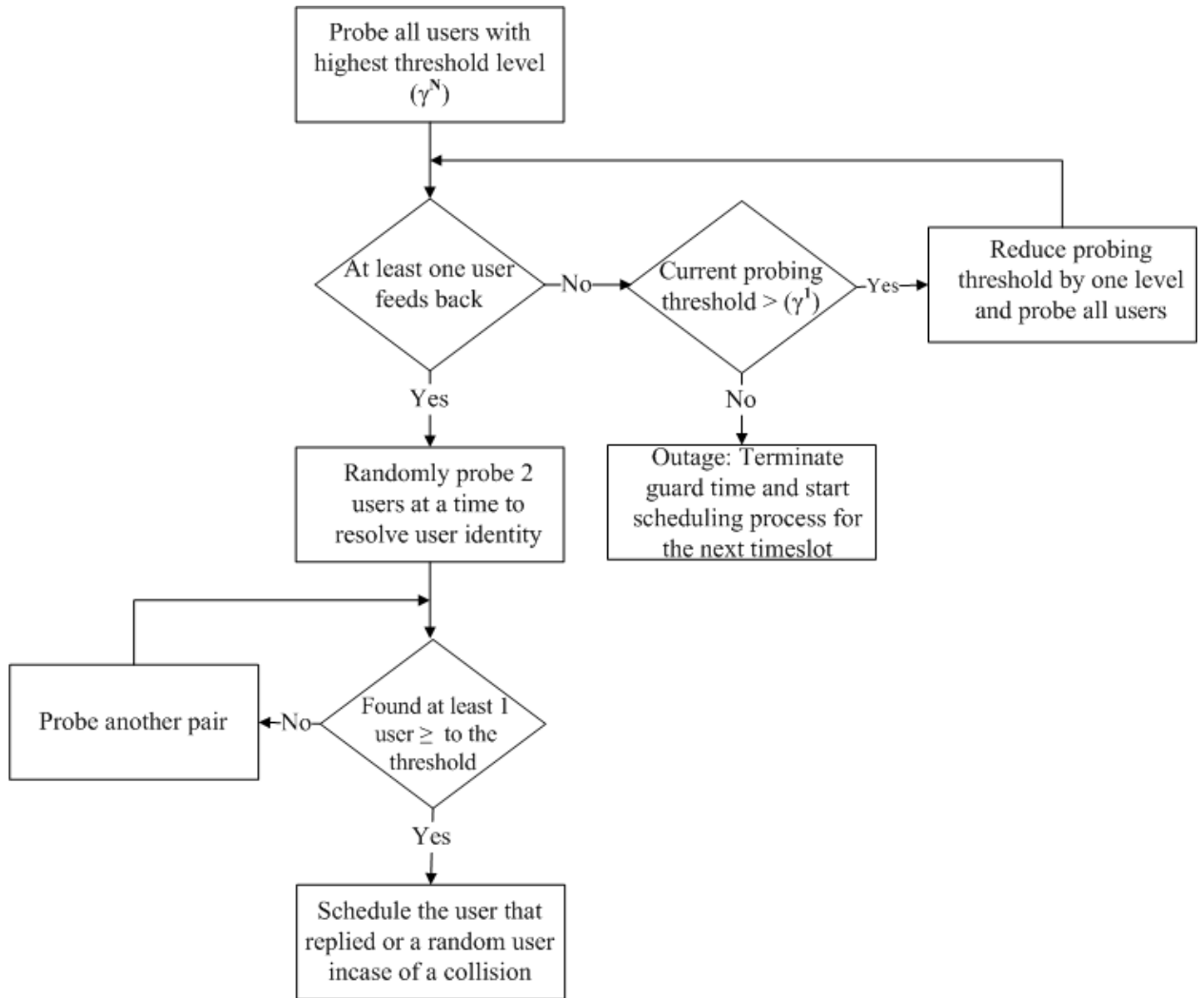


Figure 3.2: General flowchart describing the binary feedback algorithm.

terms:

(i) The first term (\overline{F}_1) measures (counts) the feedback overhead contributed by the query mode in the first stage. Mathematically, we can express \overline{F}_1 as:

$$\begin{aligned} \overline{F}_1 &= \sum_{l=2}^N l \cdot \sum_{i=1}^K \binom{K}{i} \left[F_\gamma(\gamma_{th}^{(N+2-l)}) - F_\gamma(\gamma_{th}^{(N+1-l)}) \right]^i \cdot \left[F_\gamma(\gamma_{th}^{(N+1-l)}) \right]^{K-i} \\ &+ \left(1 - F_\gamma(\gamma_{th}^{(N)})^K \right) + N \cdot \left[F_\gamma(\gamma_{th}^{(1)}) \right]^K \end{aligned} \quad (3.3)$$

where

$$F_\gamma(\gamma) = \left(1 - e^{-\frac{\gamma}{\gamma}} \right)$$

is the cumulative distribution function (CDF). From the above equation, we see that the query overhead can vary from 1 to N . If a user is found within the first query attempt, the overhead is 1. If no user is found within any of the query attempts (threshold levels), the query overhead is N as shown in the second and third terms of Equation (3.3). The first term captures the probability of finding at least one user in the second to $(N - 1)$ tries.

(ii) The second term (\overline{F}_2) results from the search mode. In this case, we have to find the probability of finding at least one user with an SNR above the threshold.

\overline{F}_2 can be expressed as

$$\begin{aligned} \bar{F}_2 &= \sum_{n=1}^{N-1} \left(\sum_{i=1}^{\nu} i \cdot P_n \cdot \left[F_{\gamma}(\gamma_{th}^{(n)}) \right]^{2i-2} \right) \cdot \left[F_{\gamma}(\gamma_{th}^{(n+1)}) \right]^K \\ &+ \sum_{i=1}^{\nu} i \cdot P_N \cdot \left[F_{\gamma}(\gamma_{th}^{(N)}) \right]^{2i-2} \end{aligned} \quad (3.4)$$

where

$$\nu = \begin{cases} \frac{K}{2} & \text{if } K \text{ is even} \\ \lfloor \frac{K}{2} \rfloor + 1 & \text{if } K \text{ is odd} \end{cases} \quad (3.5)$$

and P_n is the probability that at least one of the two selected random users has an SNR above or equal to the threshold $\gamma_{th}^{(n)}$. P_n is given by

$$P_n = \left(1 - \left[F_{\gamma}(\gamma_{th}^{(n)}) \right]^2 \right). \quad (3.6)$$

Therefore the average feedback load in this case is

$$\text{AFL} = \bar{F}_1 + \bar{F}_2. \quad (3.7)$$

3.4.2 System Capacity

Similar to [8], we consider discrete rates in this study. To find the average system capacity (C_{sys}), we first need to evaluate the average spectral efficiency (R). The average spectral efficiency is the average transmitted data rate per unit bandwidth (bits/sec/Hz) for a specific power and target bit error rate requirement. Since our

scheme always tries to select the user with the best channel conditions, it is similar to DSMUDiv [8] in terms of average spectral efficiency. R can be expressed as:

$$\begin{aligned}
R(K) &= b_0 \left([F_\gamma \gamma_{th}^{(1)}]^K \right) \\
&+ \sum_{i=1}^{N-1} b_i \left([F_\gamma \gamma_{th}^{(i+1)}]^K - [F_\gamma \gamma_{th}^{(i)}]^K \right) \\
&+ b_N \left([1 - F_\gamma \gamma_{th}^{(N)}]^K \right)
\end{aligned} \tag{3.8}$$

where b_0 is the no-transmission mode (zero) and ($b_i = \log_2 M_i$) is the number of bits per constellation.

The above expression (3.8) assumed negligible feedback traffic and did not consider it. Practically, this feedback traffic is an added overhead and must be considered. This feedback payload is simply the number of feedback information bits per probe, and it is expressed as the average feedback load multiplied by the number of bits used for feedback ($\text{AFL} \cdot \log_2 N$). Considering feedback traffic degradation, we define the average system capacity as [bits/channel use]. Accordingly, the average system capacity is expressed as:

$$C_{sys} = R(K) - \frac{\text{AFL} \cdot \log_2 N}{S} \tag{3.9}$$

where S is the number of symbols transmitted in the data transmission time-slot.

The last term in (3.9) expresses the feedback degradation caused by the feedback traffic.

3.4.3 Scheduling Delay

This section considers the impact of scheduling delay on the overall system performance. It is desired to have minimum scheduling delay to improve the overall system throughput, as this delay is part of the system resource and it must be conserved. Scheduling delay depends on the guard time duration. The guard time consists of multiple mini-slots which are used by the users to acknowledge if they are within a particular threshold level. It should be noted that the guard time duration is the time duration of the consumed mini-slots (idle counted) in the first stage (query) and the second stage (search) until a user is scheduled. Guard time measurement requires practical system parameters such as mini-slot duration, probing time, the total time-slot duration, etc. In this study, we assume that the mini-slot duration (probing time - t_p) and the time-slot duration (T_d) are 154 μsec and 50 $msec$ respectively. These parameters are based on an IEEE 802.11 based system [41] and [42] (see Table 4.1). The average guard time τ_g is expressed as:

$$\tau_g = AFL \cdot t_p \tag{3.10}$$

Table 3.2: List of used parameters

Parameter	Value
N	4 modulations
K	30 users
t_p	154 μ sec (based on [41] and [42])
T_d	5 and 50 msec (based on [41] and [42])

3.4.4 System Throughput

To get a better insight into the overall system throughput (ASTH), we consider here the effect of the guard time on the system performance. We define the average system throughput as the amount of transmitted bits per unit time (bits/sec/Hz). In (3.8), we considered the amount of transmitted bits without considering the guard time. Considering the guard time effect, we express the average system throughput as:

$$\text{ASTH} = R(K) \cdot \left(\frac{T_d - \tau_g}{T_d} \right) \quad (3.11)$$

3.5 Simulation and Numerical Results

In this section, we analyze our scheduling scheme, and we compare it with the optimal selective scheme and the DSMUDiv scheme. Figure 3.3 compares the average spectral efficiency of the proposed algorithm with both schemes. It is seen that the three algorithms have the same performance in terms of average spectral efficiency.

This stems from the fact that our scheme starts querying users with the highest threshold level and then descends to the minimum level, thus always scheduling the best user. Figure 3.4 shows the average normalized feedback load compared with the average SNR of the users. As expected, our scheme contributes to a major feedback reduction when compared with the other two schemes. This drop is due to the contribution of the query mode which optimizes the search mode probing threshold. The second stage also plays a part in the feedback reduction process, by probing two users instead of one at each mini slot duration. Furthermore, the use of a binary contention-based feedback channel, and the use of multiple thresholds, also contribute to this feedback load drop.

Feedback load reduction is highly noticeable in the region of low-to-mid average SNR. However, at high average SNR, DSMUDiv scheme performs slightly better than our scheme. This slight increase is due to the overhead caused by the first stage, as it is possible to find a user with SNR above or equal to the threshold with just a few probes without the need for threshold optimization. Figures 3.3 and 3.4 show simulation and numerical results for the average spectral efficiency and the normalized average feedback load respectively. Figures 3.5 and 3.7 show the average amount of channel state information payload on the feedback channel for the three schemes. These figures consider the amount of feedback bits (feedback rate) used to represent the feedback quantized value (for the optimal and DSMUDiv scheme) and the acknowledgement for our scheme. We ignore any other overhead, as it will

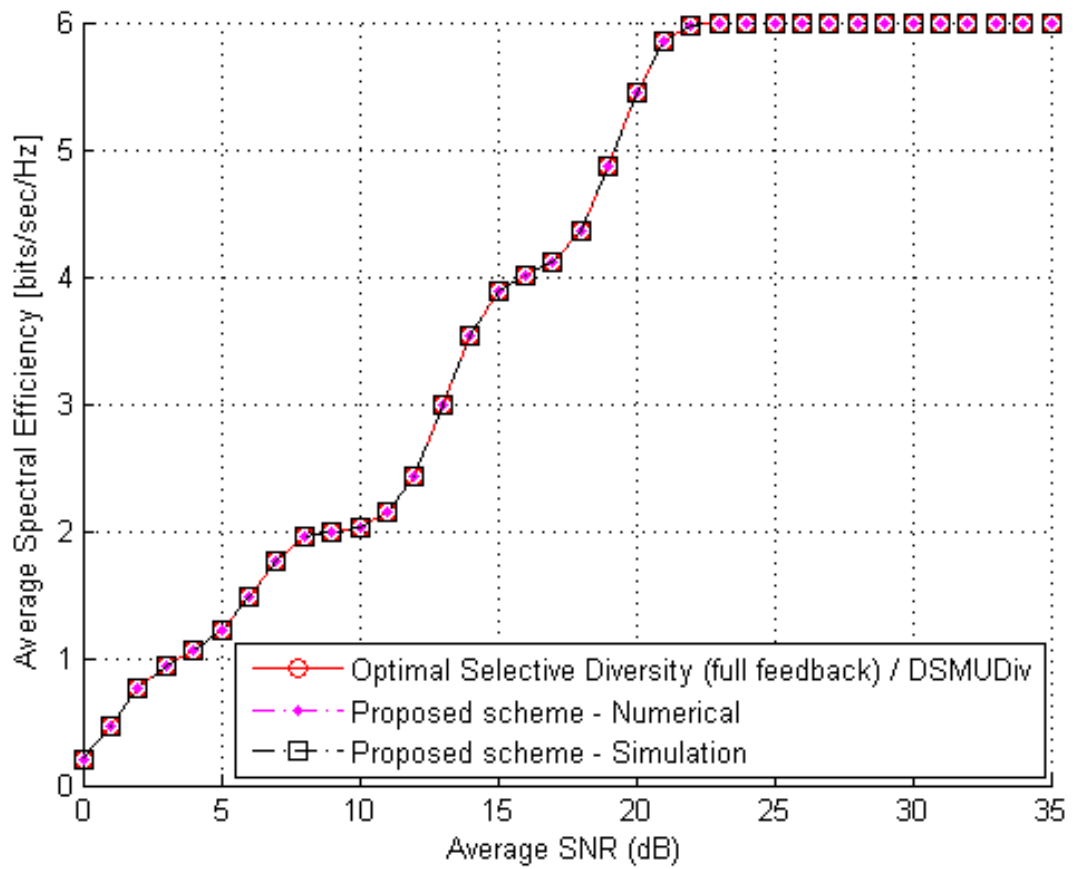


Figure 3.3: Average spectral efficiency for 30 users.

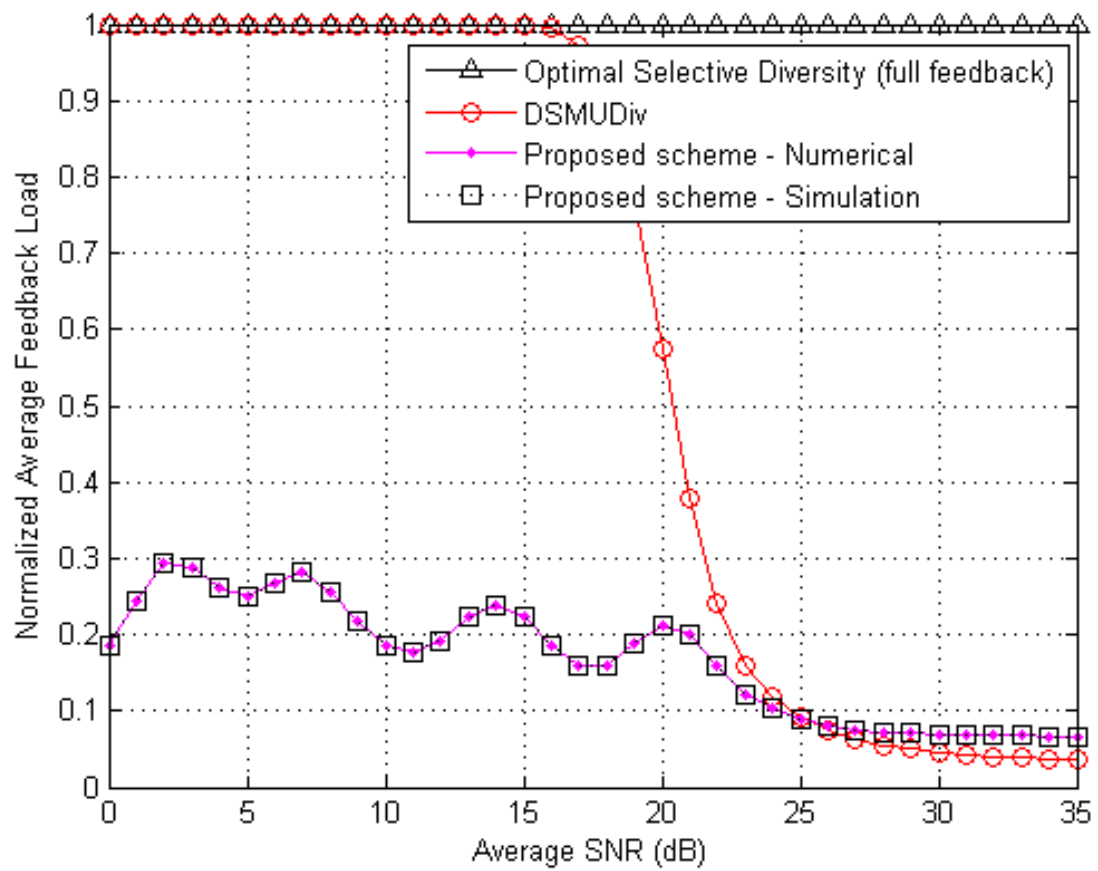


Figure 3.4: Normalized average feedback load for 30 users.

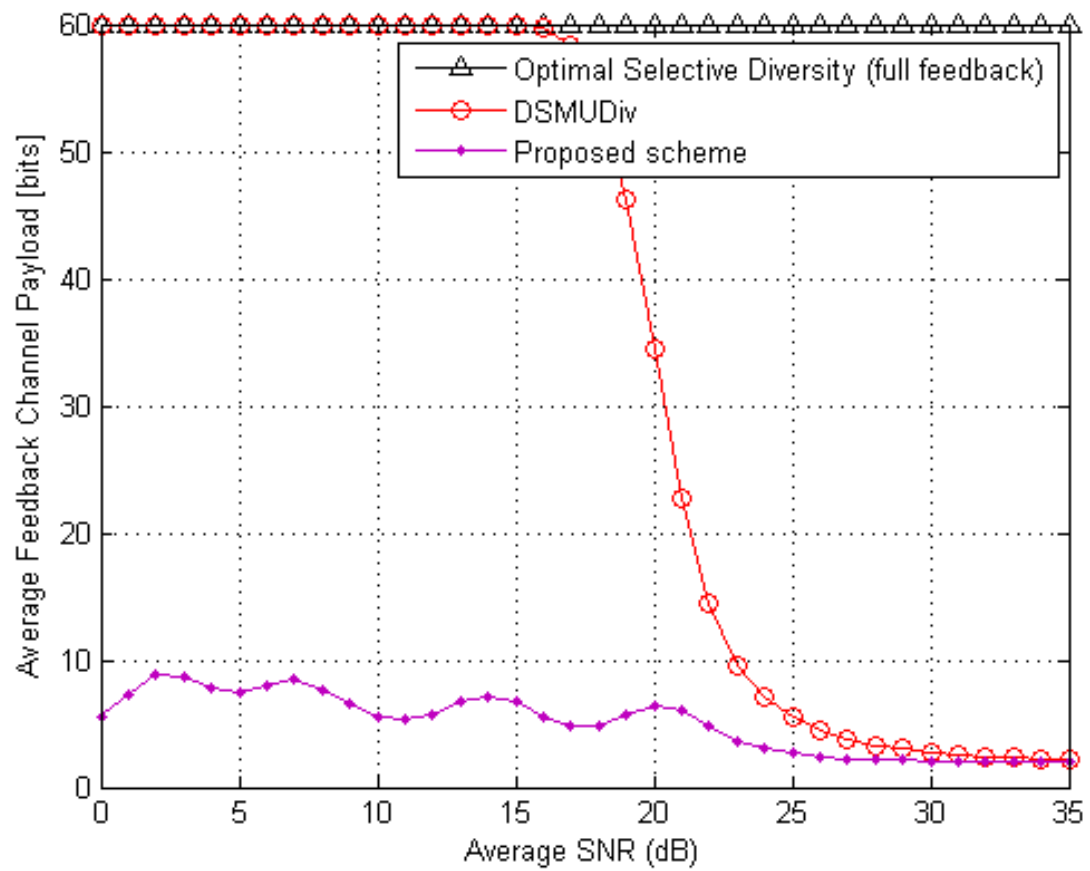


Figure 3.5: Average feedback channel payload for 30 users.

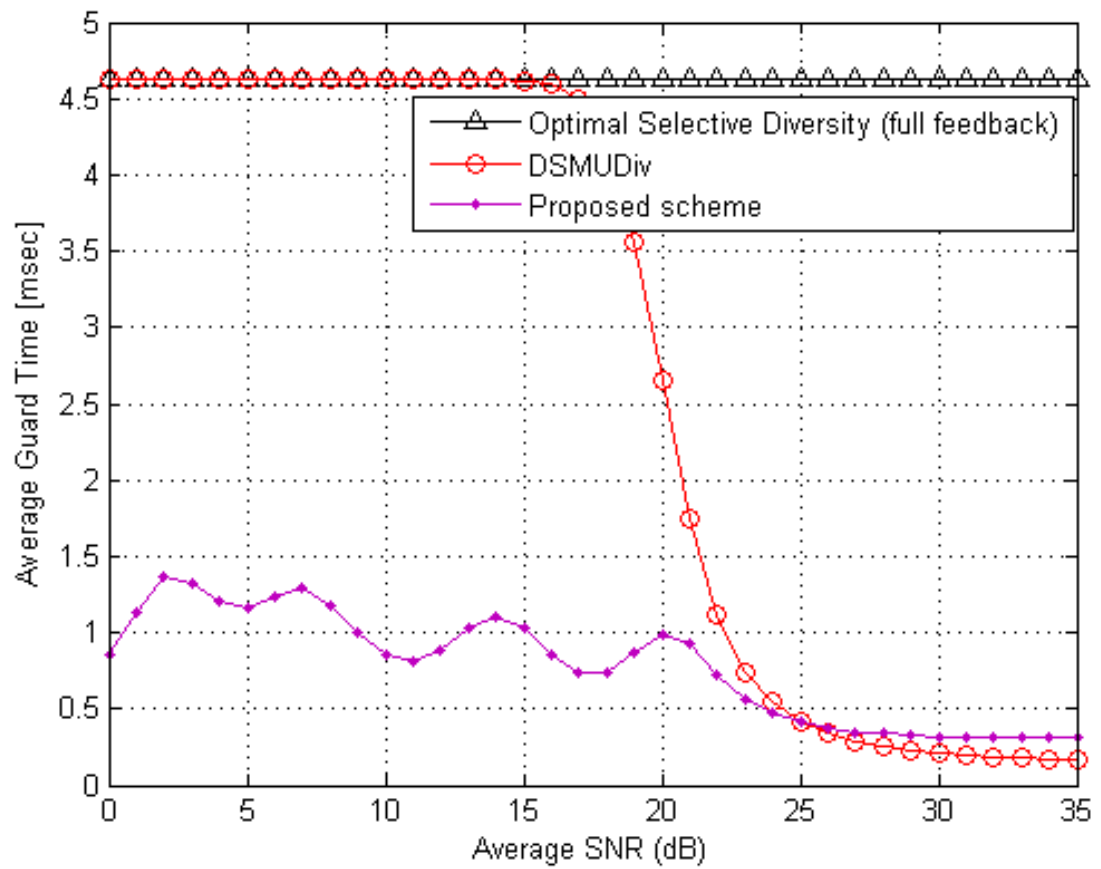


Figure 3.6: Average guard time (msec) for 30 users.

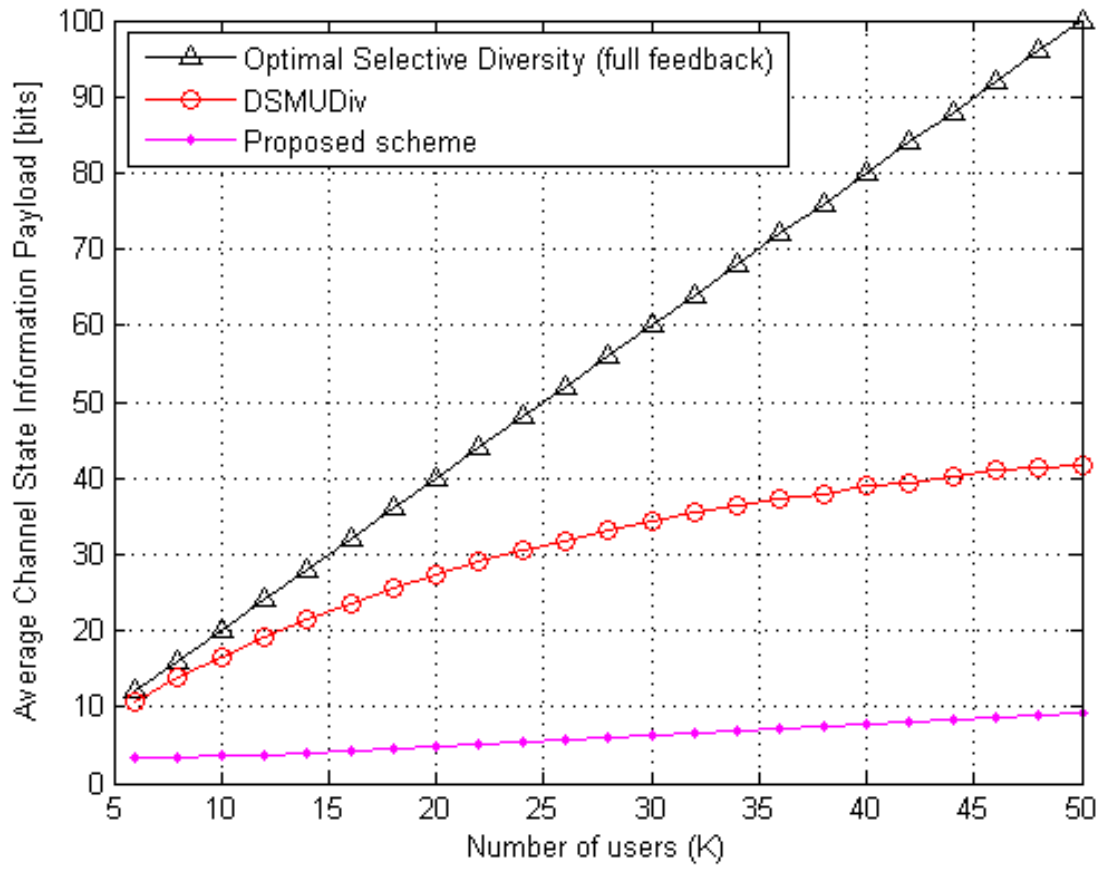


Figure 3.7: Average feedback channel payload (bits). $\tilde{\gamma} = 20$ dB.

be the same for all schemes. As seen in Figures 3.6 and 3.8, our 1-bit scheme has the minimum feedback traffic load and thus minimum guard time requirements. Figures 3.7 and 3.8 show that our scheme enjoys the minimum guard time and payload traffic requirements compared to the other two schemes when the number of users increases. This is due to the low feedback load/rate requirements of our scheme.

In Figure 3.3, the effect of feedback rate was not considered. When calculating the average system capacity, we take into account the feedback degradation which results from the channel state information traffic on the feedback channel. Even though the three schemes had the same average spectral efficiency, they do not have the same average system capacity, as shown in Figures 3.9 and 3.10.

The optimal selective scheme has the worst average system capacity due to the high amount of required feedback information. Our scheme has the best average system capacity when compared to the other two schemes. Figure 3.10 shows that the average system capacity improves as the amount of the user's data traffic increases. Finally, in Figure 3.11, we compare the overall average system throughput of our scheme with the DSMUDiv scheme. Here, we consider the guard time on the overall system performance. As shown, our 1-bit scheme still has the best performance in both short and long time-slot durations. However, both schemes suffer some capacity loss in short time-slot durations. This is due to the increase in guard time and feedback traffic as the number of users increases. The loss in our scheme

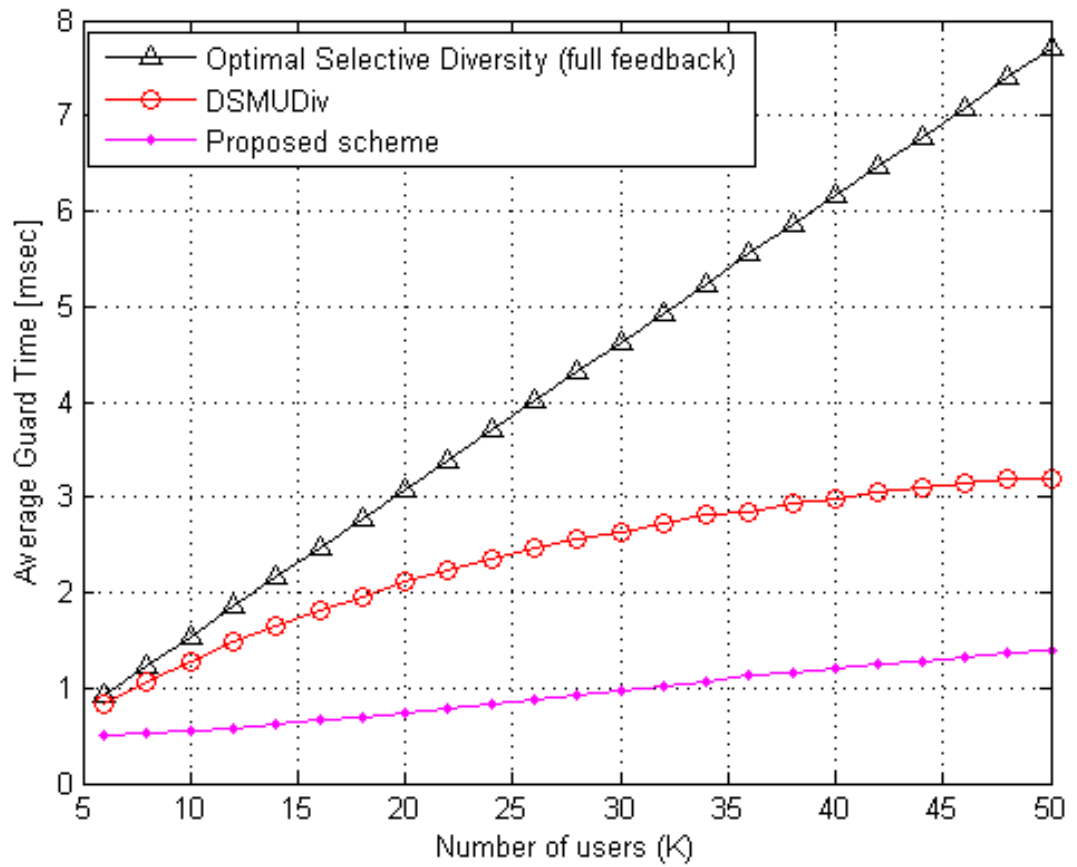


Figure 3.8: Average guard time (msec). $\tilde{\gamma} = 20$ dB.

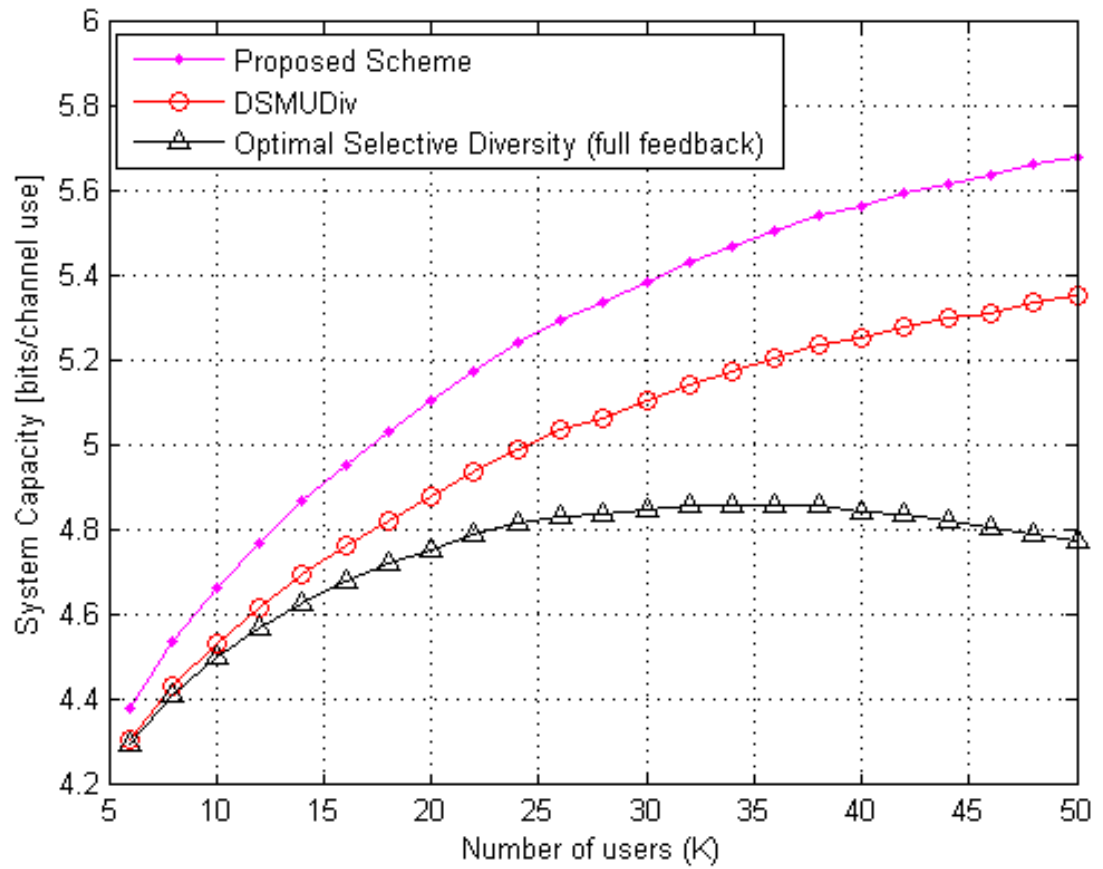


Figure 3.9: System capacity (bits/channel use) for 100 transmitted symbols with $\tilde{\gamma} = 20$ dB.

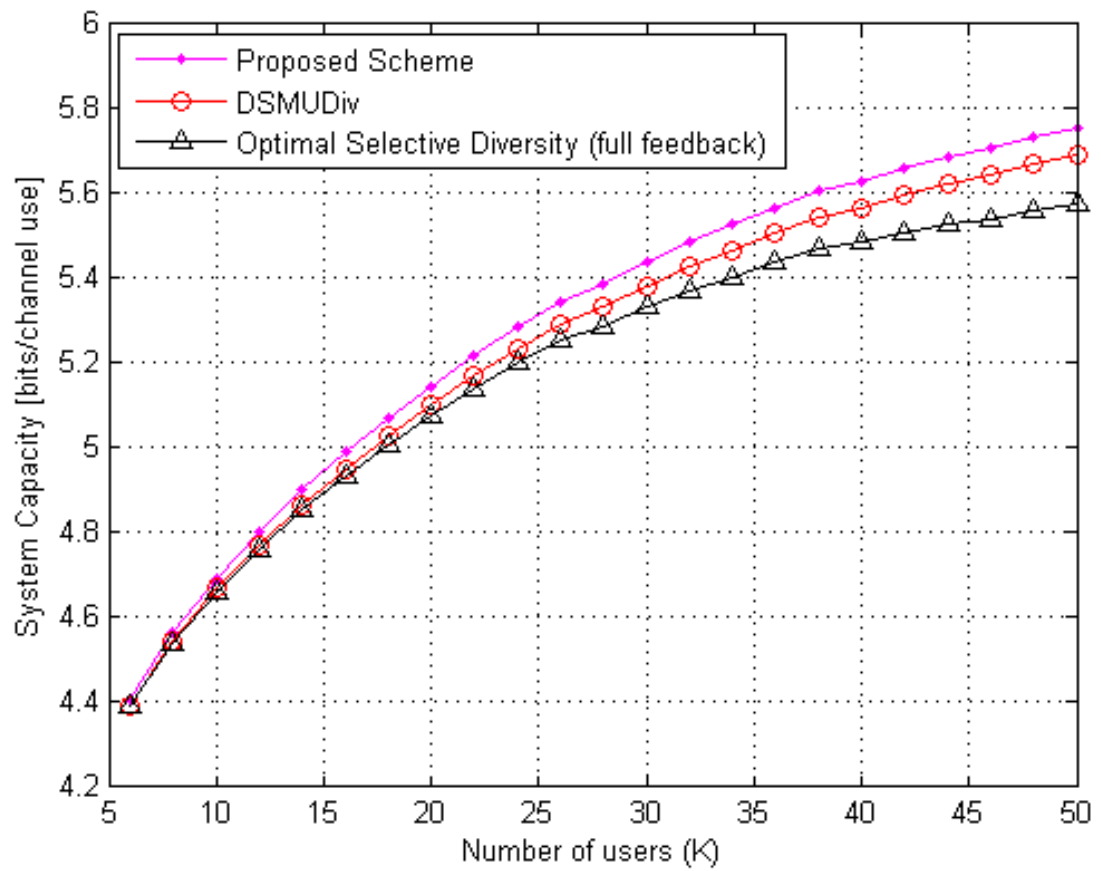


Figure 3.10: System capacity (bits/channel use) for 500 transmitted symbols with $\tilde{\gamma} = 20$ dB.

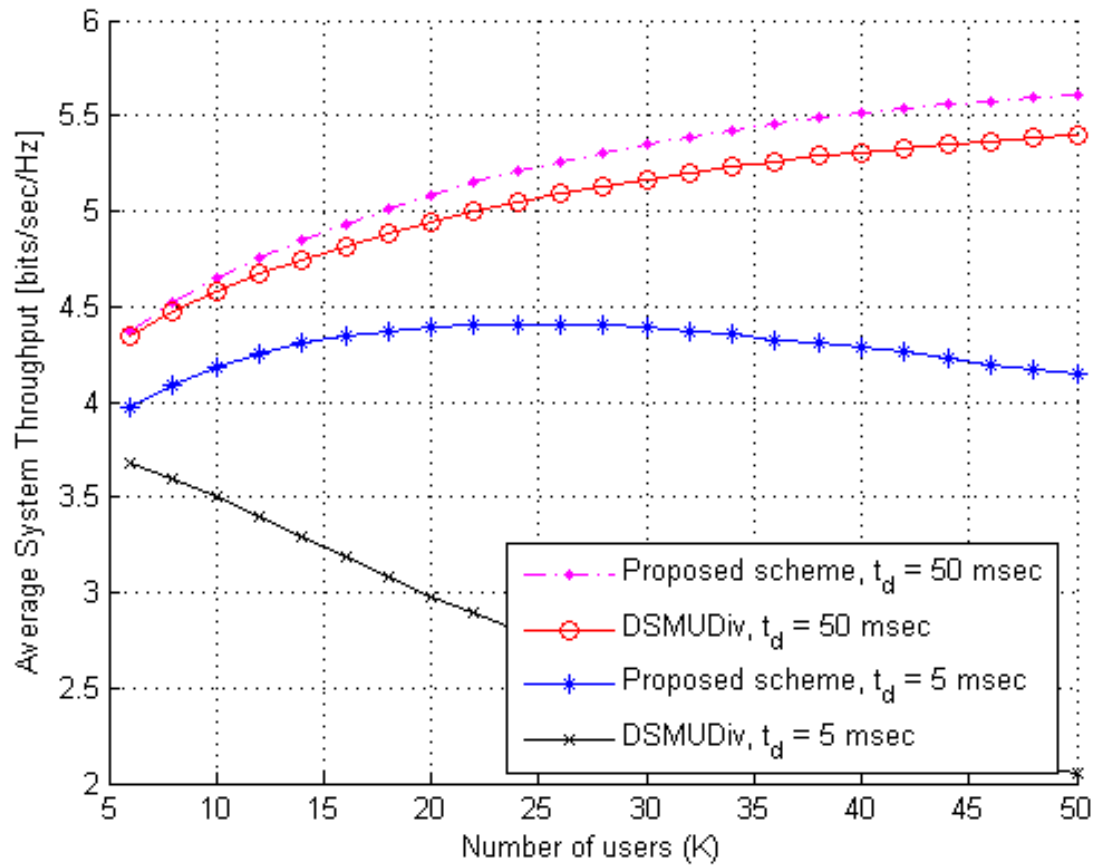


Figure 3.11: Average system throughput with time-slot durations of 5 and 50 msec. $\tilde{\gamma} = 20$ dB.

is negligible when compared to the loss incurred in the optimal scheme (not shown) and in the DSMUDiv scheme.

3.6 Conclusion

In this chapter we proposed a two-stage binary scheduling scheme that required only 1-bit feedback information per user. In the first stage, the scheduler scans all users with all of the threshold levels to estimate the threshold level on which the best user would probably be. Once at least one user is found, the scheduler goes into the second stage, where it randomly probes two users at a time and allows them to contend for one mini-slot. To reduce the feedback load, our scheme used a contention-based binary feedback channel with multiple probing threshold levels. It also employed a dual probing mechanism for probing the users. After deriving closed form expressions for the feedback load, we analyzed the performance of our scheme under slow Rayleigh fading assumption. Furthermore, when compared to the DSMUDiv scheme and the optimal selective diversity scheme, we show that our scheme reduces the feedback load, while improving the overall average system throughput and capacity with no loss in the average spectral efficiency.

Chapter 4

Opportunistic Scheduling in Multi-Carrier Systems

4.1 Introduction

Multi-carrier systems are emerging in new technologies such as xDSL, WLAN, OFDM and next generation mobile networks. The basic idea of multicarrier systems is to divide the transmitted bitstream into many different substreams and to send these substreams over many different subchannels. The number of substreams is chosen to ensure that each subchannel has a bandwidth less than the coherence bandwidth of the channel, so that the subchannels experience relatively flat fading.

In multi-carrier multiuser systems, opportunistic scheduling grants the channel resource to the user with the best channel conditions. To achieve this, the scheduler

requests CSI from all users for all carriers. The fed back CSI is considered as an overhead, and it will eventually become too high when the number of users and carriers increases. This feedback consumes a significant amount of limited system resources which could be used for data transmission. Furthermore, given a fixed coherence interval, if the guard time is not probably designed, feedback overhead could dominate the coherence time and thus lead to outdated CSI. These facts motivate researchers to propose new algorithms that reduce the feedback load while at the same time maintaining a satisfying QoS.

In this chapter, we employ a more practical model, and we consider a discrete-rate multi-carrier polling based system. Given a target BER, the system employing N modulation schemes uses N probing thresholds and allows each user to feedback quantized CSI indicating its' supported modulation level. The main dilemma in most of the proposed feedback reduction schemes is the threshold choice. A low threshold value would allow more resource allocation to users, but it would result in a capacity hit. A high threshold value improves the system capacity with increased feedback load. In our scheme, we try to maximize the system spectral efficiency, while reducing the feedback load to a minimum by using adaptive threshold levels. We present closed-form expressions for the average feedback load and average spectral efficiency. We also consider other systems parameters such as probability of access, system throughput and scheduling delay.

The remainder of this chapter is organized as follows. After the introduction,

the system model is introduced in section 4.2, followed by the proposed scheduling algorithms in section 4.3. Section 4.4 and 4.5 present the performance analysis and some numerical examples respectively. Finally, we conclude our study in section 4.6.

4.2 System Model

We consider the downlink of a multiuser multi-carrier system, as shown in Figure 4.1. Scheduling is organized on a slot basis, assuming a single interference-free cell with one access point (AP). Considering a flash-OFDM system [40] as shown in Figure 4.2, the total bandwidth is divided into L subchannels, each with S_c out of N_c subcarriers. It is assumed that the subcarriers in each subchannel have the same fading envelope and that the subchannels hold their state for the duration of one time-slot. We also assume that the fading is independent between the time-slots and that the fading coefficients of all users are independent and identically distributed (*i.i.d.*). We consider flat Rayleigh fading in this case.

A probing-based system is considered where probing of the users occurs during the guard time duration. Assuming a target bit error probability (BEP_o), the probing thresholds are given by [8] (see Table 3.1) as

$$\begin{aligned}\gamma_{th}^{(1)} &= [\text{erfc}^{-1}(2 \cdot \text{BEP}_o)]^2, \\ \gamma_{th}^{(n)} &= -\frac{2}{3}(2^n - 1) \ln(5 \cdot \text{BEP}_o); \quad n = 2, 3, \dots, N \\ \gamma_{th}^{(N+1)} &= +\infty,\end{aligned}$$

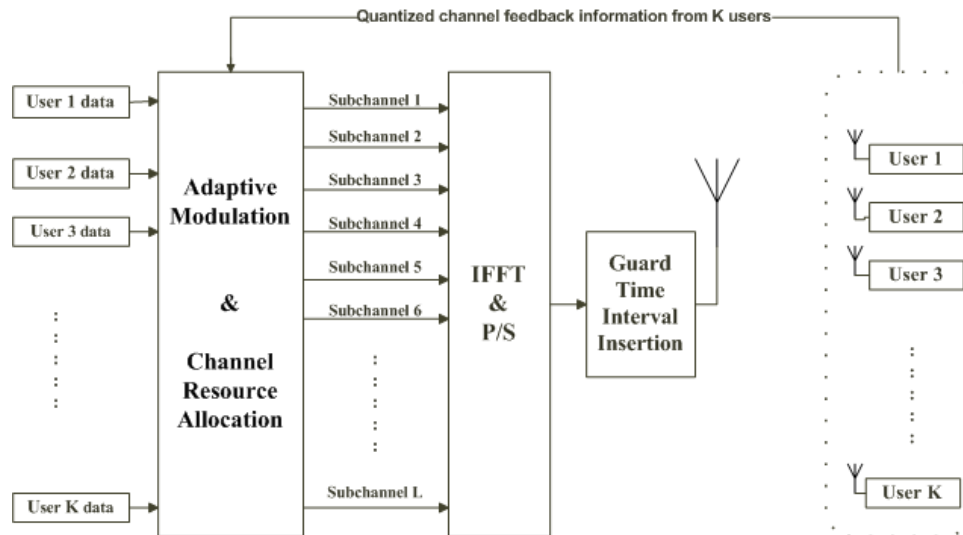


Figure 4.1: System model of the downlink multi-carrier system.

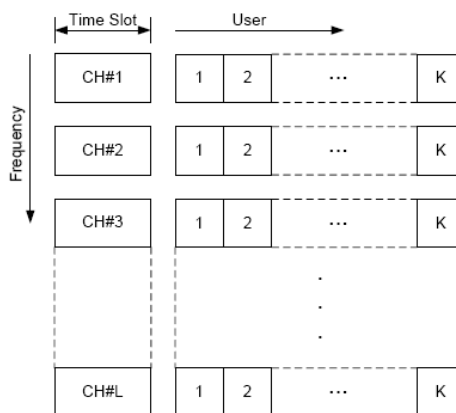


Figure 4.2: System model with L parallel subchannels and K users.

where $\text{erfc}^{-1}(\cdot)$ denotes the inverse error function. The transmission rate of each user is based on the user's channel condition, i.e. the fed back quantized value. For instance, if γ_{kl} is the estimated SNR of the k th user at the l th subchannel, then the quantized value (q_{kl}) fed back by the k th user for the l th subchannel would be [17]

$$q_{kl} = \begin{cases} q^{(0)} & \text{if } \gamma_{kl} < \gamma_{th}^{(1)} \quad (\text{outage}) \\ q^{(n)} & \text{if } \gamma_{th}^{(n)} \leq \gamma_{kl} < \gamma_{th}^{(n+1)} \\ q^{(N)} & \text{if } \gamma_{kl} \geq \gamma_{th}^{(N)} \end{cases} \quad (4.1)$$

Assuming perfect channel estimation at the receiver and an error-free feedback channel, we employ uncoded adaptive discrete rate multilevel quadrature amplitude modulation (M-QAM) similar to [17]. The modulation level which is used to modulate the k th user depends on the value of q_{kl} in (4.1). Thus the k th user can have one of the following modulation levels

$$M_k = \begin{cases} M_0 & \text{if } q^{(0)} \text{ is fed back (outage)} \\ M_n & \text{if } q^{(n)} \text{ is fed back} \end{cases} \quad (4.2)$$

where $M_0 = 1$ indicates that the user is in a deep fade and M_n , $0 \leq n \leq N$, is the modulation level corresponding to a constellation size of 2^n .

As shown, our proposed schemes utilize the different modulation levels which are already available in most of today's systems (e.g. IEEE 802.11, Flash OFDM, IEEE 802.16e), to obtain the probing thresholds. This is an added advantage which permits easy deployment of our schemes in most modern systems. The list of para-

meters used in our analysis are shown below in Table 4.1.

4.3 Scheduling Algorithms

4.3.1 Algorithm 1

The main objective of our algorithms is to reduce as much feedback load as possible while maintaining a QoS criterion with minimum service delay. The system employed is a probing-based system, where users are probed with a threshold value and the first user that is found with an instantaneous SNR above or equal to the threshold is scheduled. Intuitively, one can choose the lowest threshold value as the probing threshold. This would reduce the feedback load to a minimum at the expense of a significant amount of spectral efficiency loss. Setting a high threshold value will lead to improved spectral efficiency, but at the expense of high feedback load and increased guard time. Keeping those two points in mind, we propose an

Table 4.1: List of used parameters

Parameter	Value
N	4 modulations
N_c	52 subcarriers
K	13 and 26 users
L	13 subchannels
t_p	154 μ sec (based on [41] and [42])
T_d	5 and 50 msec (based on [41] and [42])

algorithm that adaptively changes the probing threshold. The probing threshold always starts with the minimum threshold level, and it increases gradually towards the maximum level. Whenever a user is scheduled a subchannel, the guard time is terminated, and that user does not participate in any scheduling process unless all the remaining subchannels have been scheduled. This will lead to some loss in capacity due to multiuser diversity loss.

A general flowchart of our algorithm is presented in Figure 4.3. The algorithm description is as follows:

1. There are N modulation levels, K users and L subchannels, where $K > L$.
2. Each user knows the N modulation levels and the L subchannels.
3. The AP lists all users in a list and starts the channel allocation process.
4. For the first subchannel, the AP starts probing users with the first threshold level $\gamma_{th}^{(1)}$.
5. If a user is found with an instantaneous SNR above or equal to the threshold, the user is scheduled, and the user's quantized SNR $q^{(n)}$ value is increased by one level, i.e. $q^{(m=n+1)}$, and it is set as the probing threshold for the next subchannel.
6. If no user is found, then the AP knows that all users are in a deep fade and does not schedule any user for the first subchannel.

7. For the second subchannel, the AP starts probing the remaining unscheduled users with $\gamma_{th}^{(m)}$ if the first subchannel was scheduled or $\gamma_{th}^{(1)}$ if the first subchannel was not scheduled.
8. If a user is found with instantaneous SNR above or equal to the threshold, the user is scheduled and its quantized fed-back value is stored, increased by one level, and set as the threshold value for the next subchannel.
9. If no user is found, the AP schedules the best unscheduled user, which has the highest instantaneous SNR above $\gamma_{th}^{(1)}$, and stores its quantized value $q_{max}^{(n)}$ and then increases the scheduled user's quantized value by one level, i.e. the probing threshold for the next subchannel would be $\gamma_{th}^{(m)} = q_{max}^{(m=n+1)}$.
10. Whenever a user is scheduled a subchannel, the scheduled user is removed from the list.
11. This scenario is repeated until all subchannels are scheduled.

4.3.2 Algorithm 2

In this section, we slightly modify the above proposed algorithm for further feedback reductions. Mainly, we make a slight modification to point number 9 in the algorithm description of Algorithm 1. Basically, when the expected user is not found, algorithm 1 selects the user with the highest quantized SNR, elevates the quantized SNR by

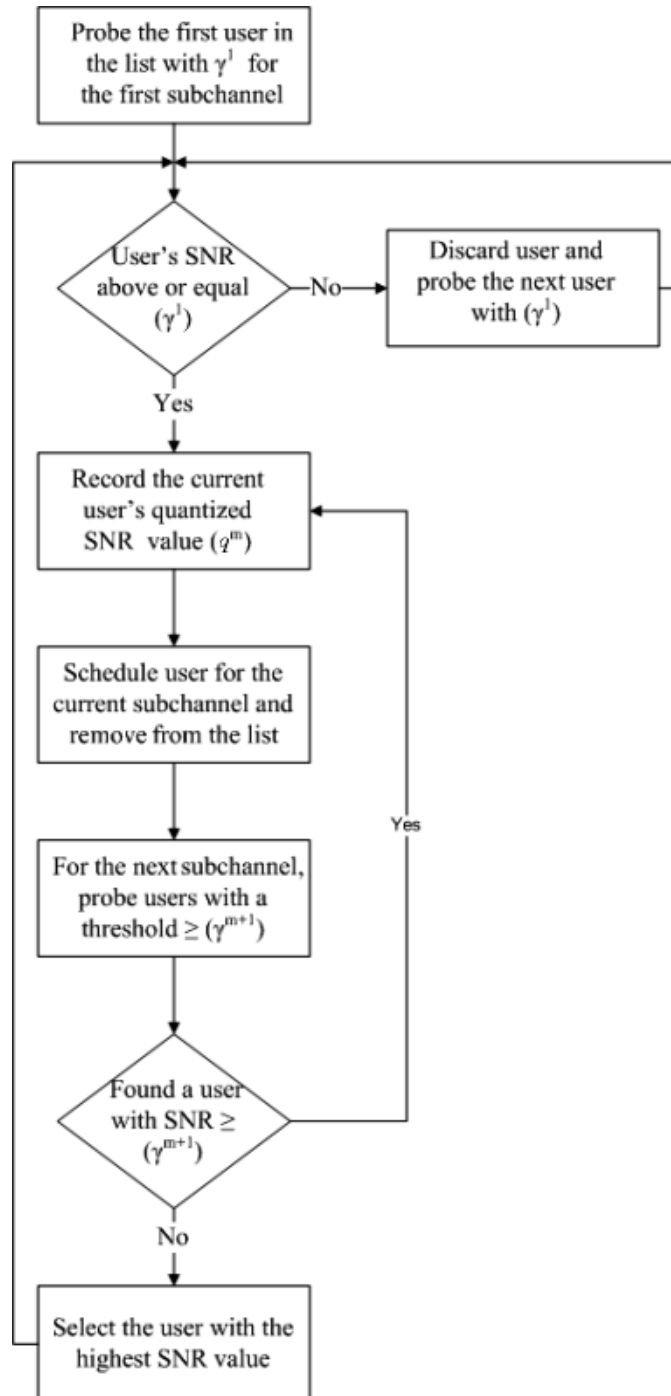


Figure 4.3: Flowchart describing Algorithm 1.

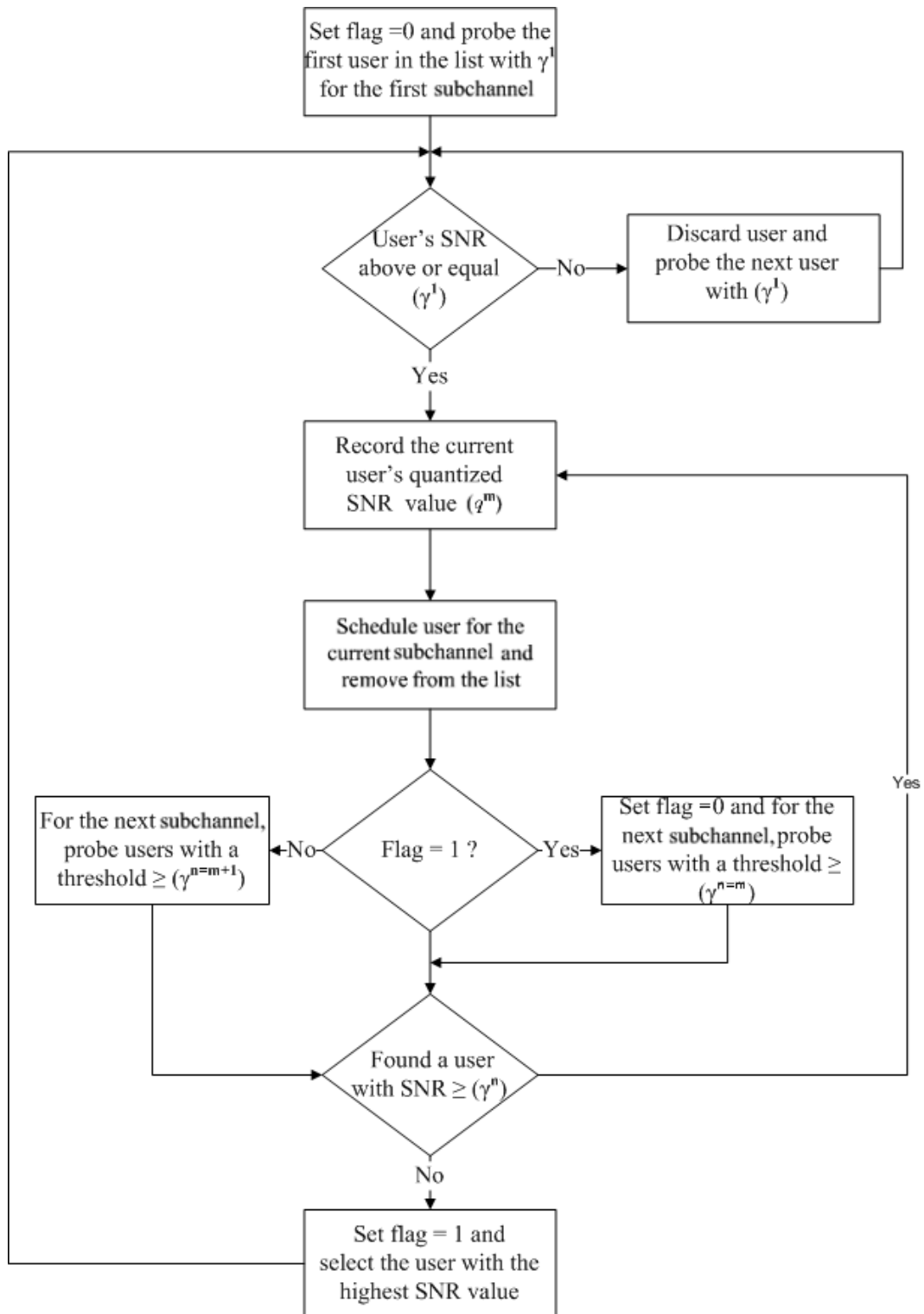


Figure 4.4: Flowchart describing Algorithm 2.

one level, and sets it as the probing threshold for the next subchannel. In algorithm 2, no elevation is performed (we do not increase the threshold by one level), i.e. instead of setting $q_{max}^{(m=n+1)}$ as the probing threshold for the next subchannel, we set $q_{max}^{(m=n)}$, where n is the fed back quantized value.

In Algorithm 1, the maximum fed back quantization level was increased by one level to make sure that the best user is always selected and to compensate for the loss incurred in the first subchannel. However, if all users had low average SNR, then this scheme would eventually lead to full probing to make a scheduling decision, as Algorithm 1 always probes the next subchannel with a threshold value higher than the instantaneous SNR of the highest user (or average SNR of the users). This problem is resolved by the modification applied here to algorithm 1. When the scheduler does not find any user within the broadcasted threshold, the probing threshold for the next subchannel is reduced and equated to the value of the highest quantized fed-back value.



Figure 4.5: Multi-carrier system time-slot model with 3 subchannels and 3 users.

Table 4.2: Probing thresholds and the corresponding fed back quantized values for the 3 algorithms approximated over 500 iterations. $K=26$, $L= 13$, $\text{BEP}_o = 10^{-4}$ and $\tilde{\gamma} = 15$ dB.

Subchannel index	DSMUDiv-EEA		Algo. 1		Algo. 2	
	Probing $\gamma_{th}^{(n)}$ (dB)	Q. F.B (dB)	Probing $\gamma_{th}^{(n)}$ (dB)	Q. F.B (dB)	Probing $\gamma_{th}^{(n)}$ (dB)	Q. F.B (dB)
1	$\gamma_{th}^{(4)}$	$q^{(3)}$	$\gamma_{th}^{(1)}$	$q^{(1)}$	$\gamma_{th}^{(1)}$	$q^{(1)}$
2	$\gamma_{th}^{(4)}$	$q^{(3)}$	$\gamma_{th}^{(2)}$	$q^{(2)}$	$\gamma_{th}^{(2)}$	$q^{(2)}$
3	$\gamma_{th}^{(4)}$	$q^{(3)}$	$\gamma_{th}^{(3)}$	$q^{(3)}$	$\gamma_{th}^{(3)}$	$q^{(3)}$
4	$\gamma_{th}^{(4)}$	$q^{(3)}$	$\gamma_{th}^{(4)}$	$q^{(3)}$	$\gamma_{th}^{(4)}$	$q^{(3)}$
5	$\gamma_{th}^{(4)}$	$q^{(3)}$	$\gamma_{th}^{(4)}$	$q^{(3)}$	$\gamma_{th}^{(3)}$	$q^{(3)}$
6	$\gamma_{th}^{(4)}$	$q^{(3)}$	$\gamma_{th}^{(4)}$	$q^{(3)}$	$\gamma_{th}^{(4)}$	$q^{(3)}$
7	$\gamma_{th}^{(4)}$	$q^{(3)}$	$\gamma_{th}^{(4)}$	$q^{(3)}$	$\gamma_{th}^{(3)}$	$q^{(3)}$
8	$\gamma_{th}^{(4)}$	$q^{(3)}$	$\gamma_{th}^{(4)}$	$q^{(3)}$	$\gamma_{th}^{(4)}$	$q^{(3)}$
9	$\gamma_{th}^{(4)}$	$q^{(3)}$	$\gamma_{th}^{(4)}$	$q^{(3)}$	$\gamma_{th}^{(3)}$	$q^{(3)}$
10	$\gamma_{th}^{(4)}$	$q^{(3)}$	$\gamma_{th}^{(4)}$	$q^{(3)}$	$\gamma_{th}^{(4)}$	$q^{(3)}$
11	$\gamma_{th}^{(4)}$	$q^{(3)}$	$\gamma_{th}^{(4)}$	$q^{(3)}$	$\gamma_{th}^{(3)}$	$q^{(2)}$
12	$\gamma_{th}^{(4)}$	$q^{(2)}$	$\gamma_{th}^{(4)}$	$q^{(2)}$	$\gamma_{th}^{(2)}$	$q^{(2)}$
13	$\gamma_{th}^{(4)}$	$q^{(2)}$	$\gamma_{th}^{(3)}$	$q^{(2)}$	$\gamma_{th}^{(3)}$	$q^{(2)}$

To have a better insight, we refer to Table 4.2. The table illustrates the probing thresholds and the quantized feedback for the three algorithms (DSMUDiv-EEA, Algorithm 1 and Algorithm 2). As shown, in low to mid average SNR, the DSMUDiv-EEA scheme requires full probing to make a scheduling decision. The scheme always probes with the highest threshold, and thus it results in full probing. Our Algorithm 1 reduced the need for full probing by starting from the minimum threshold value and then increasing towards the highest. Nevertheless, as seen in Table 4.2, Algorithm 1 required full probing from subchannel 4 to schedule all the remaining subchannels. The modified version of Algorithm 1 (Algorithm 2) reduces the need for full probing, by relaxing the probing threshold when the expected user is not found, as seen in subchannel 5 in Table 4.2.

A general flowchart for Algorithm 2 is shown in Figure 4.4. Figure 4.5 gives a general example for the time-slot model employed in our algorithms with three subchannels and three users. The time-slot is divided into a data transmission time (shaded) and a guard time period used to collect the user's feedback information. The data transmission time is fixed but the number of guard time mini-slots is variable and depends on the probability of finding a user. In this example, the number of mini-slots shows that a maximum of 3, 2 and 1 mini-slot are required for the first, second and third subchannels respectively. As noticed, each subchannel requires a single probe to find a user. The first subchannel requires 1 mini-slot to find a user but has to wait 2 mini-slots for the remaining subchannels to be

scheduled prior to its data transmission. As observed, the guard time period is reduced at each subchannel until all subchannels have been scheduled. This reduces the scheduling guard time period for the next unscheduled subchannels and thus the overall transmission latency.

4.4 Performance Analysis

In this section, we mathematically analyze the proposed scheduling algorithms. We look at the system performance in terms of feedback load, spectral efficiency, and scheduling delay. Each one of these measuring criteria will be studied.

Prior to each time-slot, a guard time-slot is dedicated for the scheduling process. We assume that the subchannels are scheduled in an ascending order ($l = 1, 2, \dots, L$). By looking at the way the scheduling scheme works, we can see that the number of users competing for subchannel assignment may vary from one subchannel to the next. This variation will depend on the previous scheduling process. With the assumption of discrete rates, a user can find himself in the no-transmission mode if his channel quality cannot support the minimum rate where reliable transmission is not possible. This situation can happen to all users at a given time-slot and subchannel. In this case, the subchannel would not be assigned to any user. Let ξ be a Bernoulli random variable with values 0 in the case a subchannel is not assigned to any user and 1 when it is assigned to a user. Given k users competing for a subchannel

assignment, the probability of ξ is:

$$P[\xi = s|k] = \begin{cases} \left(F_\gamma(\gamma_{th}^{(1)})\right)^k & \text{if } s = 0, \\ 1 - \left(F_\gamma(\gamma_{th}^{(1)})\right)^k & \text{if } s = 1, \end{cases} \quad (4.3)$$

where

$$F_\gamma(\gamma) = (1 - e^{-\frac{\gamma}{\bar{\gamma}}}) \quad (4.4)$$

the cumulative distribution function (CDF).

Let X_l denote the number of users at l th scheduling process (the l th subchannel). Then $\{X_l, l = 1, 2, \dots, L\}$ is a stochastic process that is modelled as a discrete-state Markov process (Markov chain) as shown in Figures 4.6 and 4.7. The number of states at each scheduling process is equal to the subchannel order number (l). Let $u(k, l)$ be a state at the l th scheduling process with k users in the system. In our proposed scheduling algorithms, users are probed until a user equal to or exceeding a threshold value is found. This threshold value depends on the previous scheduling process. Whenever a user is scheduled a subchannel, the user's quantized SNR value (γ_{th}^n) is increased by one level (γ_{th}^{n+1}) and it is set as a threshold value for the next subchannel. Therefore, each state is divided into substates, where each substate indicates the probing threshold value. Let \mathbf{w}_k be the set of modulation levels in each state:

$$\mathbf{w}_k = \begin{cases} \{w(n, k) = M_n : n = 1\} & ; k = K, \\ \{w(n, k) = M_n : 1 \leq n \leq N\} & ; (K - (l - 1)) < k < K, \\ \{w(n, k) = M_n : 1 < n \leq N\} & ; k = K - (l - 1). \end{cases} \quad (4.5)$$

We can see from 4.5 that the number of substates in each state will depend on the number of users in that state (k).

The transition probability from state $u(i, l)$ to state $u(j, l + 1)$ is defined as:

$$P(u(i, l), u(j, l + 1)) = \begin{cases} \lambda(i) & \text{if } j = i, \\ 1 - \lambda(i) & \text{if } j = i - 1 \end{cases} \quad (4.6)$$

where

$$\lambda(i) = \left(P[\gamma_{kl} < \gamma_{th}^{(1)}] \right)^i = \left(F_\gamma(\gamma_{th}^{(1)}) \right)^i \quad (4.7)$$

The probability of being in state $u(k, l)$ (steady-state probability) is:

$$\Gamma(k, l) = \begin{cases} \left(\lambda(K) \right)^{l-1} & \text{if } k = K, \\ \left[\left(1 - \lambda(k + 1) \right) \Gamma(k + 1, l - 1) + \lambda(k) \Gamma(k, l - 1) \right] & \text{if } (K - (l - 1)) < k < K, \\ \prod_{i=0}^{l-2} \left(1 - \lambda(K - i) \right) & \text{if } k = K - (l - 1). \end{cases} \quad (4.8)$$

In what follows, we derive the average feedback load and the average spectral efficiency for both the algorithms (Algorithm 1 and Algorithm 2.)

4.4.1 Feedback Load

The average feedback load (AFL) is defined as the average number of probes sent until the subchannel is assigned. The average feedback load conditioned on k and n is [8]:

$$\bar{F}(k, n) = \left[\sum_{i=0}^{k-1} i \left(F_{\gamma}(\gamma_{th}^{(n)}) \right)^{i-1} \left(1 - F_{\gamma}(\gamma_{th}^{(n)}) \right) \right] + k \left(F_{\gamma}(\gamma_{th}^{(n)}) \right)^{k-1}. \quad (4.9)$$

Algorithm 1

In order to derive the average feedback load and average spectral efficiency, we need to derive the steady-state probability of being in substate $w(n, k)$. The probability of being in substate $w(n, k)$ (steady-state probability) is:

$$\Phi_n(k, l) = \begin{cases} \begin{cases} \Gamma(K, l) & ; n = 1, \\ 0 & ; n > 1. \end{cases} & ; k = K, \\ \begin{cases} \Gamma(k, l - 1) \times \lambda(k) & ; n = 1, \\ \mu_n(k + 1, l - 1) & ; 2 \leq n \leq N. \end{cases} & ; (K - (l - 1)) < k < K, \\ \begin{cases} 0 & ; n = 1, \\ \mu_n(k + 1, l - 1) & ; 2 \leq n \leq N. \end{cases} & ; k = K - (l - 1), \end{cases} \quad (4.10)$$

where $\mu_n(k, l)$ is the the transition probability from substate $w_n(K, 1)$ to $w_m(k, l)$ (see Figure 4.6) and it is derived in equation 4.11.

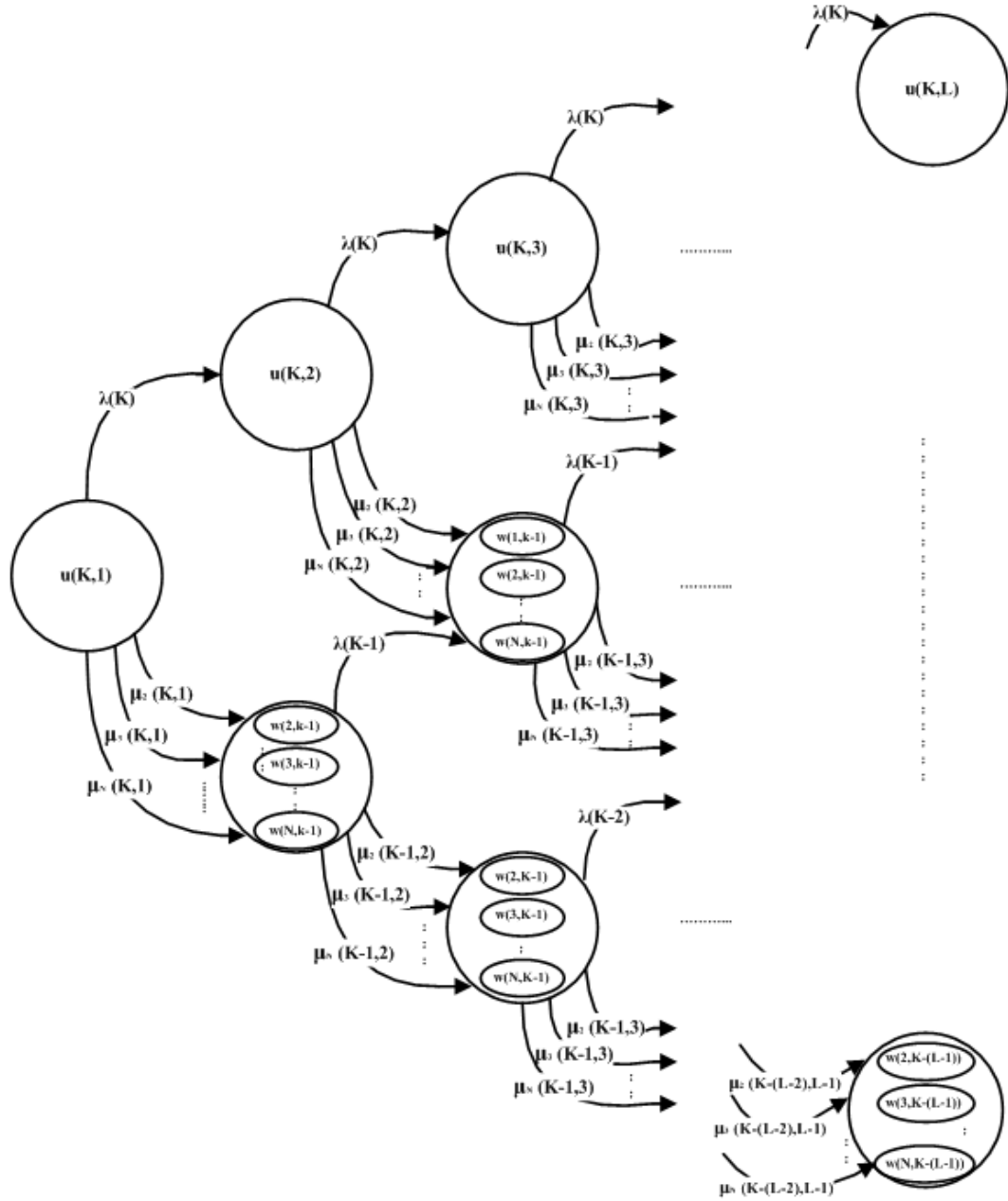


Figure 4.6: Markov chain representing algorithm 1. K is the number of users in the system and L is the number of subchannels.

As seen in Figures 4.6 and 4.7, a state does not have any substates when the number of users equals the total number of users i.e. $k = K$. In such a state, none of the users have been scheduled, and the scheduler will probe users with the minimum threshold level ($\gamma_{th}^{(1)}$) to schedule a user for the next subchannel. The second part of equation 4.10 finds the probability of being in a state which has k users, where $k > K - (l - 1)$ and less than K . Such states have N substates, which represent N threshold levels, and the substate that represents threshold level 1 ($\gamma_{th}^{(1)}$) exists here as a result of the outage which occurs in the previous state (that has the same number of users). The third part of equation 4.10 finds the probability of being in a state where the number of users $k = K - (l - 1)$ (bottom states of Figure 4.6). Such states have $(N - 1)$ substates. As observed, the substate that represents threshold level 1 does not exist here, as the algorithm always increases the probing threshold by one level, making level 2 the minimum probing threshold, as long as no outage occurs in the previous state or subchannel.

$$\mu_n(k, l) = \left\{ \begin{array}{l}
\left\{ \begin{array}{l}
0 \qquad \qquad \qquad ; n = 1, \\
\Gamma(k, l - 1) \times \sum_{i=1}^k \left(F_\gamma(\gamma_{th}^{(1)}) \right)^{i-1} \left(F_\gamma(\gamma_{th}^{(n)}) - F_\gamma(\gamma_{th}^{(n-1)}) \right) \quad ; 2 \leq n < N, \quad ; k = K, \\
\Gamma(k, l - 1) \times \sum_{i=1}^k \left(F_\gamma(\gamma_{th}^{(1)}) \right)^{i-1} \left(1 - F_\gamma(\gamma_{th}^{(n-1)}) \right) \quad ; n = N.
\end{array} \right. \\
\left\{ \begin{array}{l}
0 \quad ; n = 1, \\
\beta \quad ; 2 \leq n < N, \qquad \qquad \qquad ; (K - (l - 1)) < k < K, \\
\alpha \quad ; n = N.
\end{array} \right. \\
\left\{ \begin{array}{l}
0 \qquad \qquad \qquad ; n = 1, \\
\Gamma(k, l) \times \sum_{i=1}^k \binom{k}{i} \left(F_\gamma(\gamma_{th}^{(1)}) \right)^{k-i} \left(F_\gamma(\gamma_{th}^{(2)}) - F_\gamma(\gamma_{th}^{(1)}) \right)^i \quad ; n = 2, \quad ; k = K - (l - 1). \\
\delta \qquad \qquad \qquad ; 2 < n < N, \\
\eta \qquad \qquad \qquad ; n = N.
\end{array} \right.
\end{array} \right.$$

(4.11)

where

$$\begin{aligned}
\beta = & \left\{ \sum_{i=1}^k \binom{k}{i} \left(F_{\gamma}(\gamma_{th}^{(n-1)}) \right)^{k-i} \left(F_{\gamma}(\gamma_{th}^{(n)}) - F_{\gamma}(\gamma_{th}^{(n-1)}) \right)^i \right\} \\
& \times \left\{ \sum_{z=n}^N \mu_z(k+1, l-1) \right\} \\
& + \left\{ \sum_{y=1}^{n-2} \left[\sum_{i=1}^k \left(F_{\gamma}(\gamma_{th}^{(n-y)}) \right)^{i-1} \left(F_{\gamma}(\gamma_{th}^{(n)}) - F_{\gamma}(\gamma_{th}^{(n-1)}) \right) \right. \right. \\
& \quad \left. \left. \times \mu_{n-y}(k+1, l-1) \right] \right\} \\
& + \left\{ \sum_{i=1}^k \left(F_{\gamma}(\gamma_{th}^{(1)}) \right)^{i-1} \left(F_{\gamma}(\gamma_{th}^{(n)}) - F_{\gamma}(\gamma_{th}^{(n-1)}) \right) \right\} \\
& \times \left\{ \Gamma(k, l-1) \times \lambda(k) \right\}
\end{aligned} \tag{4.12}$$

and

$$\begin{aligned}
\alpha = & \left\{ \sum_{i=1}^k \binom{k}{i} \left(F_{\gamma}(\gamma_{th}^{(N-1)}) \right)^{k-i} \left(F_{\gamma}(\gamma_{th}^{(N)}) - F_{\gamma}(\gamma_{th}^{(N-1)}) \right)^i \right. \\
& \left. + \sum_{i=1}^k \left(F_{\gamma}(\gamma_{th}^{(N)}) \right)^{i-1} \left(1 - F_{\gamma}(\gamma_{th}^{(N)}) \right) \right\} \\
& \times \left\{ \mu_N(k+1, l-1) \right\} \\
& + \left\{ \sum_{y=1}^{N-2} \left[\sum_{i=1}^k \left(F_{\gamma}(\gamma_{th}^{(N-y)}) \right)^{i-1} \left(1 - F_{\gamma}(\gamma_{th}^{(N-1)}) \right) \right. \right. \\
& \quad \left. \left. \times \mu_{(N-y)}(k+1, l-1) \right] \right\} \\
& + \left\{ \sum_{i=1}^k \left(F_{\gamma}(\gamma_{th}^{(1)}) \right)^{i-1} \left(1 - F_{\gamma}(\gamma_{th}^{(N-1)}) \right) \right\} \\
& \times \left\{ \Gamma(k, l-1) \times \lambda(k) \right\}
\end{aligned} \tag{4.13}$$

and

$$\begin{aligned}
\delta = & \left\{ \sum_{i=1}^k \binom{k}{i} \left(F_{\gamma}(\gamma_{th}^{(n-1)}) \right)^{k-i} \left(F_{\gamma}(\gamma_{th}^{(n)}) - F_{\gamma}(\gamma_{th}^{(n-1)}) \right)^i \right\} \\
& \times \left\{ \sum_{z=n}^N \mu_z(k+1, l-1) \right\} \\
& + \left\{ \sum_{y=1}^{n-2} \left[\sum_{i=1}^k \left(F_{\gamma}(\gamma_{th}^{(n-y)}) \right)^{i-1} \left(F_{\gamma}(\gamma_{th}^{(n)}) - F_{\gamma}(\gamma_{th}^{(n-1)}) \right) \right. \right. \\
& \quad \left. \left. \times \mu_{(n-y)}(k+1, l-1) \right] \right\}
\end{aligned} \tag{4.14}$$

and

$$\begin{aligned}
\eta = & \left\{ \sum_{i=1}^k \binom{k}{i} \left(F_{\gamma}(\gamma_{th}^{(N-1)}) \right)^{k-i} \left(F_{\gamma}(\gamma_{th}^{(N)}) - F_{\gamma}(\gamma_{th}^{(N-1)}) \right)^i \right. \\
& \left. + \sum_{i=1}^k \left(F_{\gamma}(\gamma_{th}^{(N)}) \right)^{i-1} \left(1 - F_{\gamma}(\gamma_{th}^{(N)}) \right) \right\} \\
& \times \left\{ \mu_N(k+1, l-1) \right\} \\
& + \left\{ \sum_{y=1}^{N-2} \left[\sum_{i=1}^k \left(F_{\gamma}(\gamma_{th}^{(N-y)}) \right)^{i-1} \left(1 - F_{\gamma}(\gamma_{th}^{(N-1)}) \right) \right. \right. \\
& \quad \left. \left. \times \mu_{(N-y)}(k+1, l-1) \right] \right\}.
\end{aligned} \tag{4.15}$$

In equation 4.11, we find the transition probability from a state to a state or a substate to a substate. In such cases, knowledge of the transition probability is required for the current and previous states only. The first part of equation 4.11 represents the transition probability for the upper states of Figure 4.6. In such states, we need to find the transition probability from one threshold level to $(N-1)$ threshold levels, as these states have only one threshold level. The second and third part of equation 4.11 represents the transition probabilities of the mid and bottom

states of Figure 4.6. In such states, each state consists of more than one substate. This complicates the mathematical derivation of the transitional probability. The Greek symbols shown in equation 4.11 are used for simplification purposes only. The symbols β and α represent the transition probabilities from the middle states to middle states, and the symbols δ and η represent the transition probabilities from the lower states to lower states only, as shown in Figure 4.6.

Therefore, the average feedback load at the l th scheduling process is:

$$\text{AFL}^{sub}(l) = \begin{cases} \bar{F}(K, 1) & \text{if } l = 1, \\ \left\{ \begin{array}{l} \sum_{j=2}^{l-1} \sum_{n=1}^N \bar{F}(K - (j - 1), n) \Phi_n(K - (j - 1), l) \\ + \sum_{n=2}^N \bar{F}(K - (l - 1), n) \Phi_n(k, l) \\ + \bar{F}(K, 1) \Phi_1(K, l) \end{array} \right. & \text{if } 1 < l \leq L. \end{cases} \quad (4.16)$$

By normalizing over the number of subchannels we get

$$\text{AFL} = \frac{1}{L} \left[\sum_{l=1}^L \text{AFL}^{sub}(l) \right]. \quad (4.17)$$

Algorithm 2

Similarly, as in algorithm 1, here we derive the steady-state probability of being in substate $w(n, k)$, the average feedback load and the average spectral efficiency for algorithm 2. The analysis is almost the same as above, but the only difference here is that we have an extra substate in the bottom states, which represents threshold

level 1. This is shown in Figure 4.7.

The probability of being in substate (steady state probability) $w(n, k)$ is:

$$\Psi_n(k, l) = \begin{cases} \begin{cases} \Gamma(K, l) & ; n = 1, \\ 0 & ; n > 1. \end{cases} & ; k = K, \\ \begin{cases} \Gamma(k, l - 1) \times \lambda(k) & ; n = 1, \\ \mu_n(k + 1, l - 1) & ; 2 \leq n \leq N. \end{cases} & ; (K - (l - 1)) < k < K, \\ \mu_n(k + 1, l - 1) & ; k = K - (l - 1). \end{cases} \quad (4.18)$$

The transition probability is (see Figure 4.7):

$$\mu_n(k, l) = \begin{cases} \begin{cases} 0 & ; n = 1, \\ \Gamma(k, l - 1) \times \sum_{i=1}^k \left(F_{\gamma}(\gamma_{th}^{(1)}) \right)^{i-1} \left(F_{\gamma}(\gamma_{th}^{(n)}) - F_{\gamma}(\gamma_{th}^{(n-1)}) \right) & ; 2 \leq n < N, \\ \Gamma(k, l - 1) \times \sum_{i=1}^k \left(F_{\gamma}(\gamma_{th}^{(1)}) \right)^{i-1} \left(1 - F_{\gamma}(\gamma_{th}^{(n-1)}) \right) & ; n = N. \end{cases} & ; k = K, \\ \begin{cases} v & ; n = 1, \\ \beta & ; 2 \leq n < N, \\ \alpha & ; n = N. \end{cases} & ; (K - (l - 1)) < k < K, \\ \begin{cases} \delta & ; 1 \leq n < N, \\ \eta & ; n = N. \end{cases} & ; k = K - (l - 1). \end{cases} \quad (4.19)$$

where

$$\begin{aligned}
v = & \left\{ \sum_{i=1}^k \binom{k}{i} \left(F_{\gamma}(\gamma_{th}^{(n)}) \right)^{k-i} \left(F_{\gamma}(\gamma_{th}^{(n+1)}) - F_{\gamma}(\gamma_{th}^{(n)}) \right)^i \right\} \\
& \times \left\{ \sum_{z=n+1}^N \mu_z(k+1, l-1) \right\}
\end{aligned} \tag{4.20}$$

and

$$\begin{aligned}
\beta = & \left\{ \sum_{i=1}^k \binom{k}{i} \left(F_{\gamma}(\gamma_{th}^{(n)}) \right)^{k-i} \left(F_{\gamma}(\gamma_{th}^{(n+1)}) - F_{\gamma}(\gamma_{th}^{(n)}) \right)^i \right\} \\
& \times \left\{ \sum_{z=n+1}^N \mu_z(k+1, l-1) \right\} \\
& + \left\{ \sum_{y=1}^{n-1} \left[\sum_{i=1}^k \left(F_{\gamma}(\gamma_{th}^{(n-y)}) \right)^{i-1} \left(F_{\gamma}(\gamma_{th}^{(n)}) - F_{\gamma}(\gamma_{th}^{(n-1)}) \right) \right. \right. \\
& \quad \left. \left. \times \mu_{n-y}(k+1, l-1) \right] \right\} \\
& + \left\{ \sum_{i=1}^k \left(F_{\gamma}(\gamma_{th}^{(1)}) \right)^{i-1} \left(F_{\gamma}(\gamma_{th}^{(n)}) - F_{\gamma}(\gamma_{th}^{(n-1)}) \right) \right\} \\
& \times \left\{ \Gamma(k, l-1) \times \lambda(k) \right\}
\end{aligned} \tag{4.21}$$

and

$$\begin{aligned}
\alpha = & \left\{ \sum_{i=1}^k \left(F_{\gamma}(\gamma_{th}^{(N)}) \right)^{i-1} \left(1 - F_{\gamma}(\gamma_{th}^{(N)}) \right) \right\} \\
& \times \left\{ \mu_N(k+1, l-1) \right\} \\
& + \left\{ \sum_{y=1}^{N-1} \left[\sum_{i=1}^k \left(F_{\gamma}(\gamma_{th}^{(N-y)}) \right)^{i-1} \left(1 - F_{\gamma}(\gamma_{th}^{(N-1)}) \right) \right. \right. \\
& \quad \left. \left. \times \mu_{(N-y)}(k+1, l-1) \right] \right\} \\
& + \left\{ \sum_{i=1}^k \left(F_{\gamma}(\gamma_{th}^{(1)}) \right)^{i-1} \left(1 - F_{\gamma}(\gamma_{th}^{(N-1)}) \right) \right\} \\
& \times \left\{ \Gamma(k, l-1) \times \lambda(k) \right\}
\end{aligned} \tag{4.22}$$

and

$$\begin{aligned}
\delta = & \left\{ \sum_{i=1}^k \binom{k}{i} \left(F_{\gamma}(\gamma_{th}^{(n)}) \right)^{k-i} \left(F_{\gamma}(\gamma_{th}^{(n+1)}) - F_{\gamma}(\gamma_{th}^{(n)}) \right)^i \right\} \\
& \times \left\{ \sum_{z=n+1}^N \mu_z(k+1, l-1) \right\} \\
& + \left\{ \sum_{y=1}^{n-1} \left[\sum_{i=1}^k \left(F_{\gamma}(\gamma_{th}^{(n-y)}) \right)^{i-1} \left(F_{\gamma}(\gamma_{th}^{(n)}) - F_{\gamma}(\gamma_{th}^{(n-1)}) \right) \right. \right. \\
& \quad \left. \left. \times \mu_{(n-y)}(k+1, l-1) \right] \right\}.
\end{aligned} \tag{4.23}$$

and

$$\begin{aligned}
\eta = & \left\{ \sum_{i=1}^k \left(F_{\gamma}(\gamma_{th}^{(N)}) \right)^{i-1} \left(1 - F_{\gamma}(\gamma_{th}^{(N)}) \right) \right\} \\
& \times \left\{ \mu_N(k+1, l-1) \right\} \\
& + \left\{ \sum_{y=1}^{N-1} \left[\sum_{i=1}^k \left(F_{\gamma}(\gamma_{th}^{(N-y)}) \right)^{i-1} \left(1 - F_{\gamma}(\gamma_{th}^{(N-1)}) \right) \right. \right. \\
& \quad \left. \left. \times \mu_{(N-y)}(k+1, l-1) \right] \right\}.
\end{aligned} \tag{4.24}$$

The average feedback load at the l th scheduling process is:

$$\text{AFL}^{sub}(l) = \begin{cases} \bar{F}(K, 1) & \text{if } l = 1, \\ \left\{ \begin{aligned} & \sum_{j=2}^{l-1} \sum_{n=1}^N \bar{F}(K - (j-1), n) \Psi_n(K - (j-1), l) \\ & + \sum_{n=2}^N \bar{F}(K - (l-1), n) \Psi_n(k, l) \\ & + \bar{F}(K, 1) \Psi_1(K, l) \end{aligned} \right\} & \text{if } 1 < l \leq L. \end{cases} \tag{4.25}$$

Similar to Algorithm 1, by normalizing over the number of subchannels we get

$$\text{AFL} = \frac{1}{L} \left[\sum_{l=1}^L \text{AFL}^{sub}(l) \right]. \tag{4.26}$$

4.4.2 Average Spectral Efficiency

The average spectral efficiency (ASE) is defined as the average transmitted data rate per unit bandwidth in bits/sec/Hz for specified power and target error performance.

Algorithm 1

The average spectral efficiency conditioned on k and n is:

$$\begin{aligned}
\bar{R}(k, n) = & b_0 \cdot \left(F_\gamma(\gamma_{th}^{(1)}) \right)^k \\
& + \sum_{i=1}^{n-1} b_i \cdot \left[\left(F_\gamma(\gamma_{th}^{(i+1)}) \right)^k - \left(F_\gamma(\gamma_{th}^{(i)}) \right)^k \right] \\
& + \sum_{i=n}^{N-1} b_i \cdot \left[\sum_{j=1}^k \left(F_\gamma(\gamma_{th}^{(n)}) \right)^{j-1} \left(F_\gamma(\gamma_{th}^{(i+1)}) - F_\gamma(\gamma_{th}^{(i)}) \right) \right] \\
& + b_N \cdot \left[\sum_{j=1}^k \left(F_\gamma(\gamma_{th}^{(n)}) \right)^{j-1} \left(1 - F_\gamma(\gamma_{th}^{(N)}) \right) \right].
\end{aligned} \tag{4.27}$$

where $b_n = \log_2 M_n$ is the number of bits per constellation. By averaging the spectral efficiency over all users in all possible states (see Figure 4.6), we get the average spectral efficiency at the l th scheduling process ($\text{ASE}^{sub}(l)$)

$$\text{ASE}^{sub}(l) = \begin{cases} \bar{R}(K, 1) & \text{if } l = 1, \\ \left\{ \begin{aligned} & \sum_{j=2}^{l-1} \sum_{n=1}^N \bar{R}(K - (j-1), n) \Phi_n(K - (j-1), l) \\ & + \sum_{n=2}^N \bar{R}(K - (l-1), n) \Phi_n(k, l) \\ & + \bar{R}(K, 1) \Phi_1(K, l) \end{aligned} \right. & \text{if } 1 < l \leq L. \end{cases} \tag{4.28}$$

By normalizing over the number of subchannels we get

$$\text{ASE} = \frac{1}{L} \left[\sum_{l=1}^L \text{ASE}^{sub}(l) \right]. \tag{4.29}$$

Algorithm 2

Similar to Algorithm 1, by averaging the spectral efficiency over all users in all possible states (see Figure 4.7), we get the average spectral efficiency (bits/sec/Hz) conditioned on k and n at the l th scheduling process ($\text{ASE}^{sub}(l)$)

$$\text{ASE}^{sub}(l) = \begin{cases} \bar{R}(K, 1) & \text{if } l = 1, \\ \left\{ \begin{array}{l} \sum_{j=2}^{l-1} \sum_{n=1}^N \bar{R}(K - (j - 1), n) \Psi_n(K - (j - 1), l) \\ + \sum_{n=2}^N \bar{R}(K - (l - 1), n) \Psi_n(k, l) \\ + \bar{R}(K, 1) \Psi_1(K, l) \end{array} \right. & \text{if } 1 < l \leq L. \end{cases} \quad (4.30)$$

By normalizing over the number of subchannels we get

$$\text{ASE} = \frac{1}{L} \left[\sum_{l=1}^L \text{ASE}^{sub}(l) \right]. \quad (4.31)$$

4.4.3 Probability of Access

In this section, we analyze the access probability of users, assuming that all users have the same QoS requirements. At each subchannel, at most one user can be scheduled. Therefore, the probability of access per subchannel, given that k users are competing for the subchannel assignment, is [8]:

$$P_{sub}(k) = \frac{\left(1 - F_{\gamma}(\gamma_{th}^{(1)}) \right)}{k}, \quad (4.32)$$

The number of users varies at each subchannel, depending on which state it is in. Taking this into account, the probability that a user gets assigned a subchannel out of the L subchannels, which is the probability of access, is:

$$P_{access} = 1 - \prod_{l=1}^L \Lambda^{sub}(l), \quad (4.33)$$

where

$$\Lambda^{sub}(l) = \begin{cases} \Upsilon(K) & \text{if } l = 1, \\ \sum_{j=1}^l \Upsilon(K - (j - 1))\Gamma(K - (j - 1), l) & \text{if } 1 < l \leq L \end{cases} \quad (4.34)$$

and

$$\Upsilon(k) = 1 - P_{sub}(k). \quad (4.35)$$

4.4.4 Scheduling Delay

Scheduling delay is an important parameter in guard time (τ_g) analysis. Our proposed scheme is a polling-based scheduling scheme, where the users respond to the probing by feeding back their channels' qualities. This probing is performed during the guard time period, which is between the bursts. To calculate the guard time, we need real-time measurements of the probing time (τ_p) of each user. We assume that the probing time is equal for all users. The subchannel scheduling time delay, which is the time needed to schedule a subchannel, is a function of the number of probes, which is the feedback load. Therefore, the average time delay to schedule

the l th subchannel is:

$$\bar{\tau}_{sub}(l) = \begin{cases} \bar{F}(K) \cdot \tau_p & \text{if } l = 1, \\ \left(\sum_{j=1}^l \bar{F}(K - (j - 1)) \Gamma(K - (j - 1), l) \right) \cdot \tau_p & \text{if } 1 < l \leq L. \end{cases} \quad (4.36)$$

Ignoring other overhead, the average guard time is the sum of L subchannels' scheduling delays. Therefore, the average guard time is:

$$\bar{\tau}_g = \sum_{l=1}^L \bar{\tau}_{sub}(l). \quad (4.37)$$

4.4.5 Average System Throughput

The system throughput is defined as the amount of bits transmitted per unit time, where this time includes the data transmission time (T_d) and the guard time (τ_g). When deriving the ASE, we did not consider the effect of the scheduling delay. Hence, we derive the average system throughput (ASTH) by taking into account the effect of the guard time duration. The normalized average system throughput is:

$$\begin{aligned} \text{ASTH} &= \frac{1}{L} \left[\sum_{l=1}^L \left(\frac{T_d - \bar{\tau}_{sub}(l)}{T_d} \right) \text{ASE}^{sub}(l) \right] \\ &= \frac{1}{L} \left[\sum_{l=1}^L \text{ASE}^{sub}(l) - \sum_{l=1}^L \left(\frac{\bar{\tau}_{sub}(l)}{T_d} \right) \text{ASE}^{sub}(l) \right]. \end{aligned} \quad (4.38)$$

4.5 Numerical Examples

In this section, we provide some numerical examples of our proposed algorithms, and we compare the performance of the proposed schemes with the optimal DSMUDiv

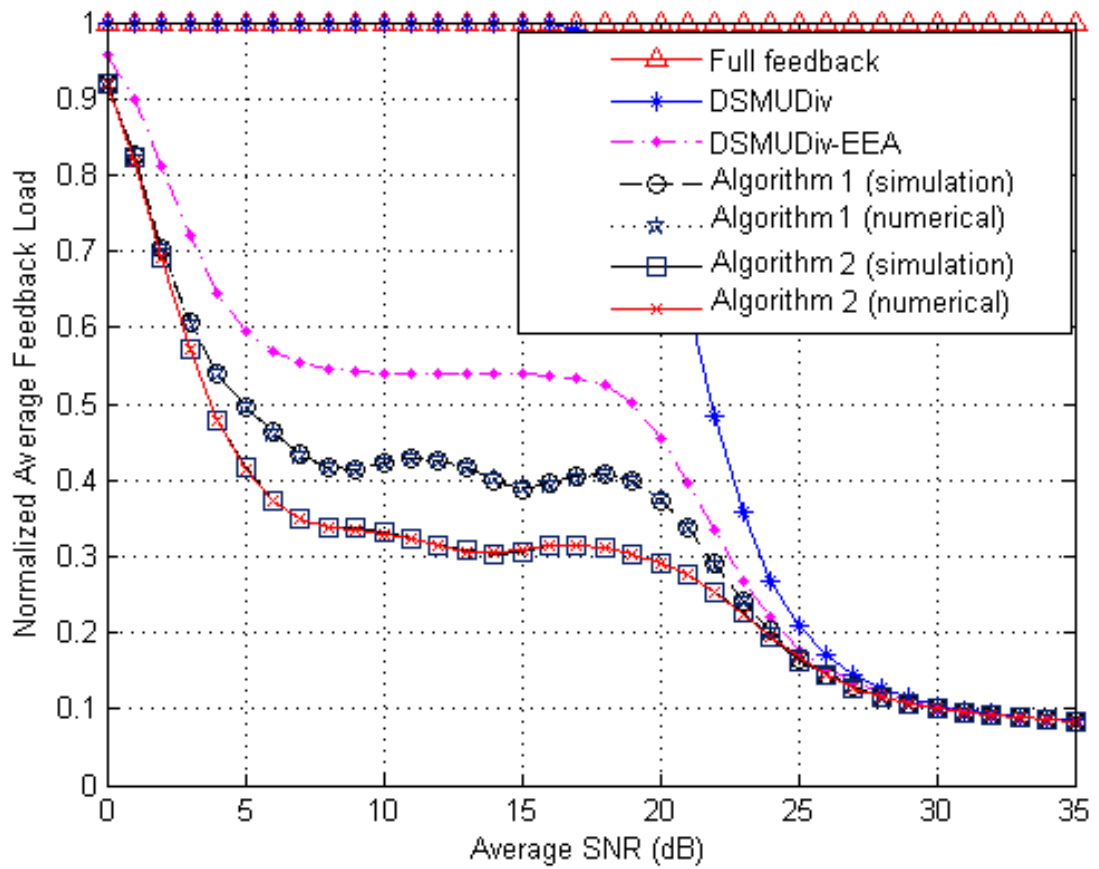


Figure 4.8: Normalized Average feedback load versus the average SNR. $K = 13$.

and DSMUDiv-EEA schemes [8] [17]. As mentioned previously, some parameters require realtime practical measurements which are out of the scope of this work. Hence, we use the parameters shown in Tables 3.1 and 4.1.

Figures 4.8 and 4.10 compare the normalized average feedback load per subchannel for all the 5 schemes. As shown, the optimal scheme has unity load, as it always requires feedback from all users to schedule the best user. The DSMUDiv scheme reduced the feedback load at high average SNR, but it still required full feedback load at low average SNR to make a scheduling decision. The DSMUDiv-EEA scheme further reduced the feedback load, as it prohibited a user competing for more than one subchannel until all subchannels were scheduled. As observed, our adaptive proposed schemes contribute to more feedback reduction, with a slight penalty loss in the average spectral efficiency, as shown in Figures 4.9 and 4.11.

Two main factors contribute to the loss shown in Figures 4.9 and 4.11. The first is due to multiuser diversity loss when scheduled users are deprived from the scheduling process. The second results from the probing threshold which starts from the minimum level and gradually increases towards the maximum level. Our adaptive threshold schemes increase the probability of finding a user with an acceptable instantaneous SNR, but not the best instantaneous SNR, thus resulting in fewer probes and less feedback load. As expected, our proposed algorithm (2) further reduces the feedback load at low to mid average SNR with slight capacity loss.

The effect of scheduling delay is shown in Figures 4.12 and 4.13. All schemes

demand higher guard time constraints as the number of users increases, and our schemes consume the minimum time when compared to the other three schemes. Other overhead is assumed to be negligible, and it is ignored here as it will be the same for the four schemes. However, special attention is required to ensure that the coherence time is not dominated by the guard time.

Figures 4.14 and 4.15 show that our proposed schemes yield the same access probability as the DSMUDiv-EEA scheme. The probability increases as the average SNR increases for both our proposed schemes and the DSMUDiv-EEA scheme, and they have better access probabilities when compared to the optimal and DSMUDiv schemes. This is due to the fact that, in each scheduling process, K users are competing for the channel resource in the optimal and DSMUDiv schemes, while every scheduled user is removed from the scheduling process in the DSMUDiv-EEA and proposed schemes.

Finally, we consider the effect of scheduling delay on the overall average system throughput as defined in section 4.4. As depicted in Figures 4.16 and 4.18, our proposed schemes give the best performance when compared to the other three schemes at low to mid average SNR when the AP is transmitting in short time-slot durations. This stems from the fact that our schemes require less average guard time to schedule users. Nevertheless, the ASTH suffers a slight loss in high average SNR regime. As mentioned before, this is partly due to multiuser diversity loss. The optimal scheme has the worst ASTH when compared to the other schemes, as

it requires more guard time to schedule the best user. Figures 4.17 and 4.19 show that our schemes have almost the same ASTH as the DSMUDiv-EEA scheme, and both schemes have ASTH below that of the optimal and DSMUDiv scheme when the AP is transmitting data by using longer time-slot durations. The reason for this is that, in long time-slot durations, the effect of guard time is negligible and the ASE dominates the ASTH. In spite of this, the gap between the ASTH of the DSMUDiv scheme and our proposed schemes shrinks as the number of users increases, as shown in Figure 4.19.

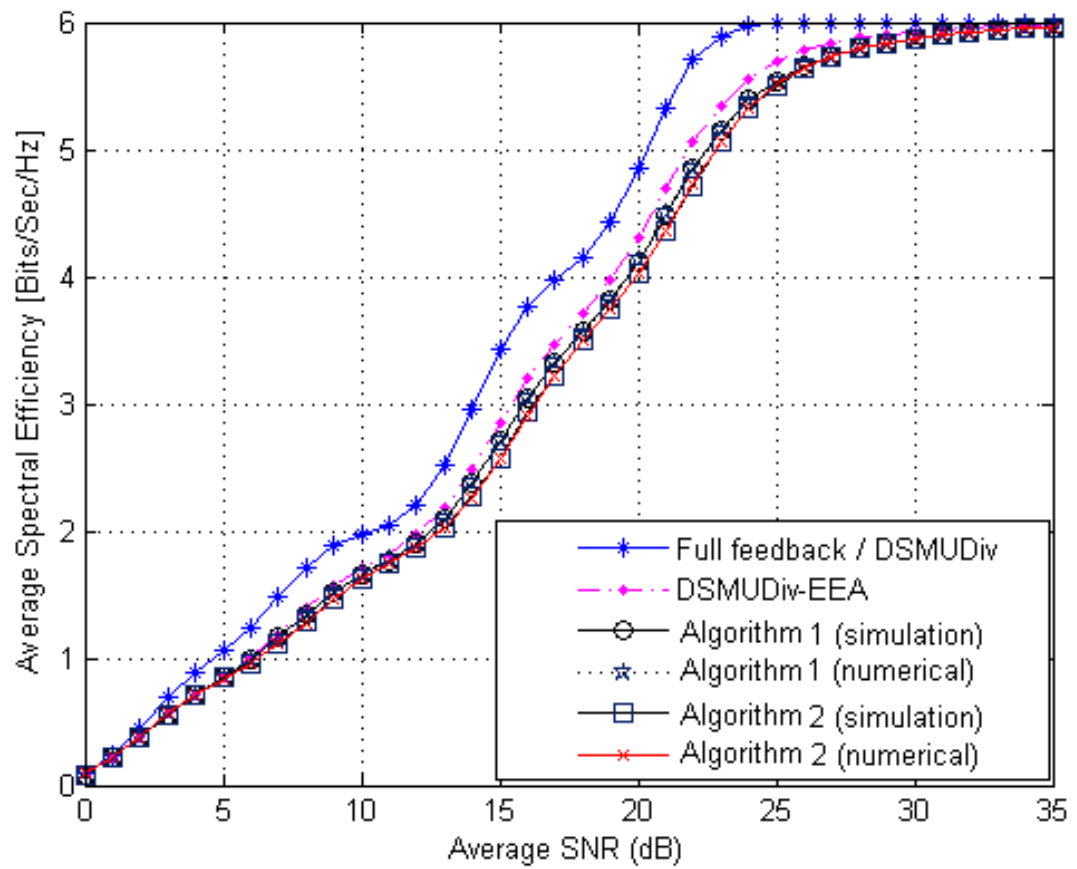


Figure 4.9: Average spectral efficiency versus the average SNR. $K = 13$.

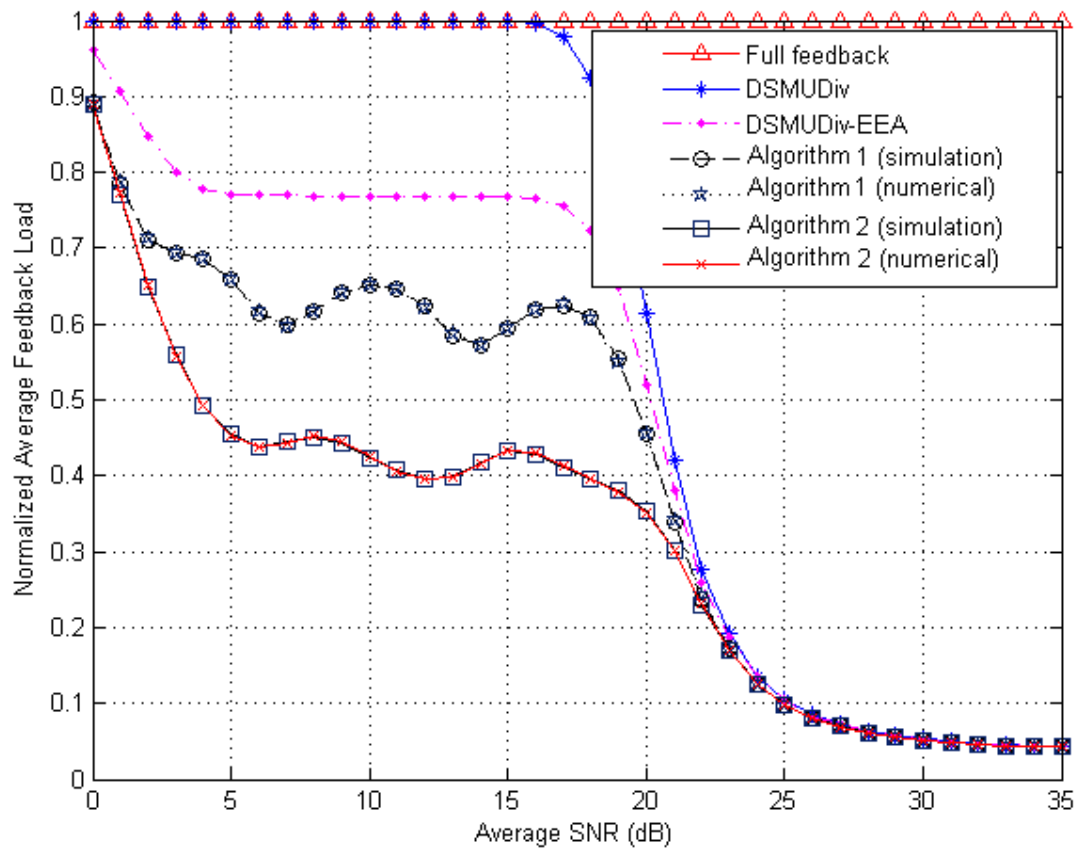


Figure 4.10: Normalized Average feedback load versus the average SNR. $K = 26$.

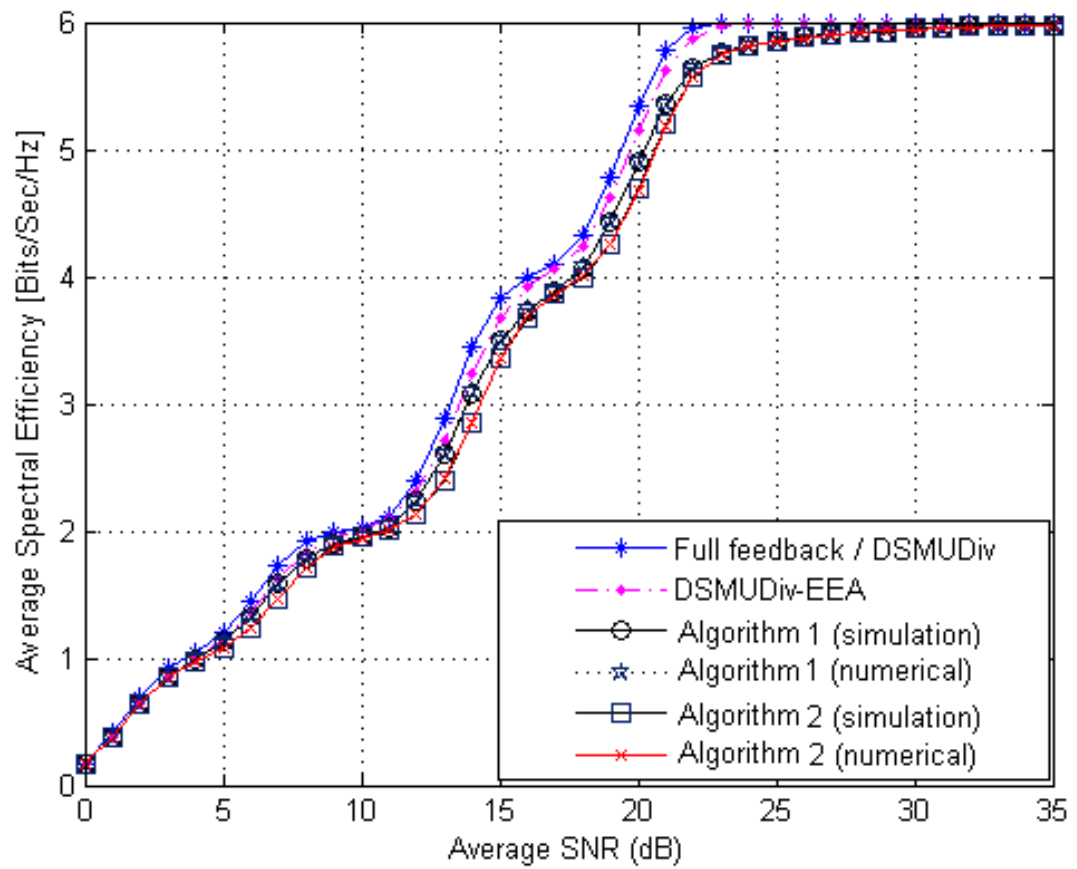


Figure 4.11: Average spectral efficiency versus the average SNR. $K = 26$.

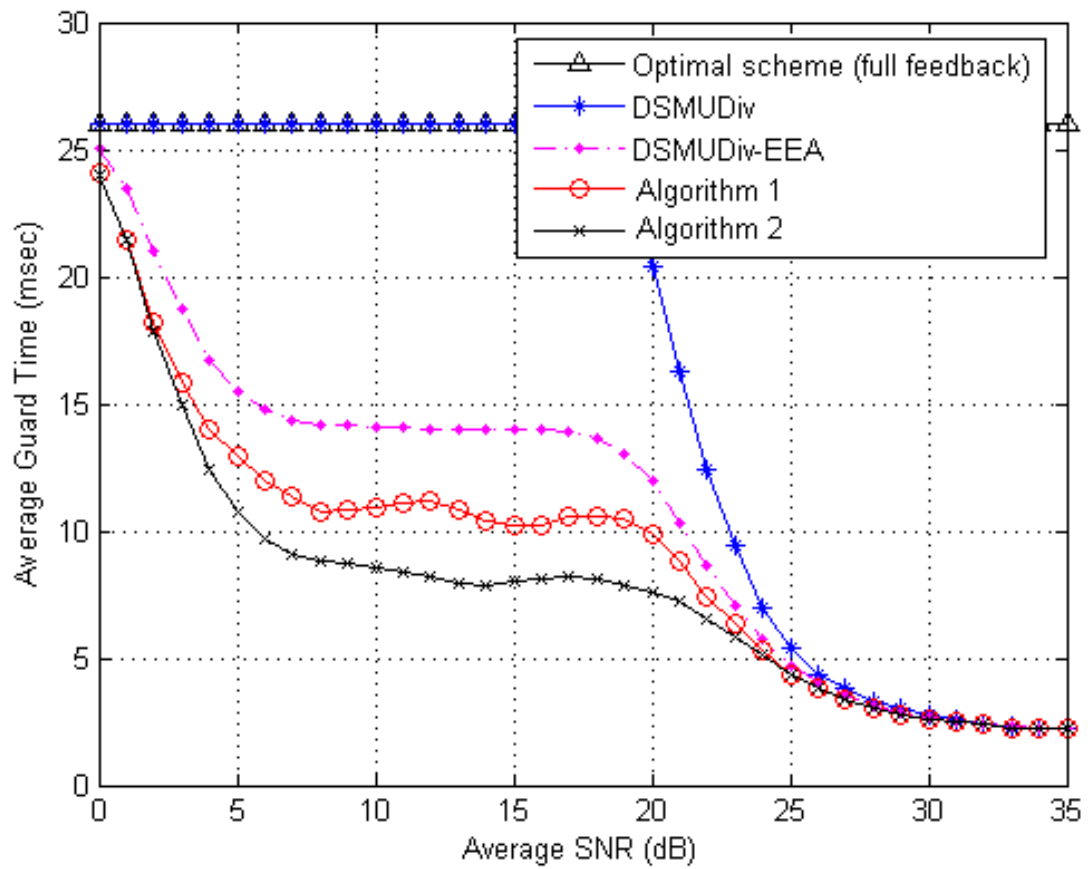


Figure 4.12: Average guard time versus the average SNR. $K = 13$.

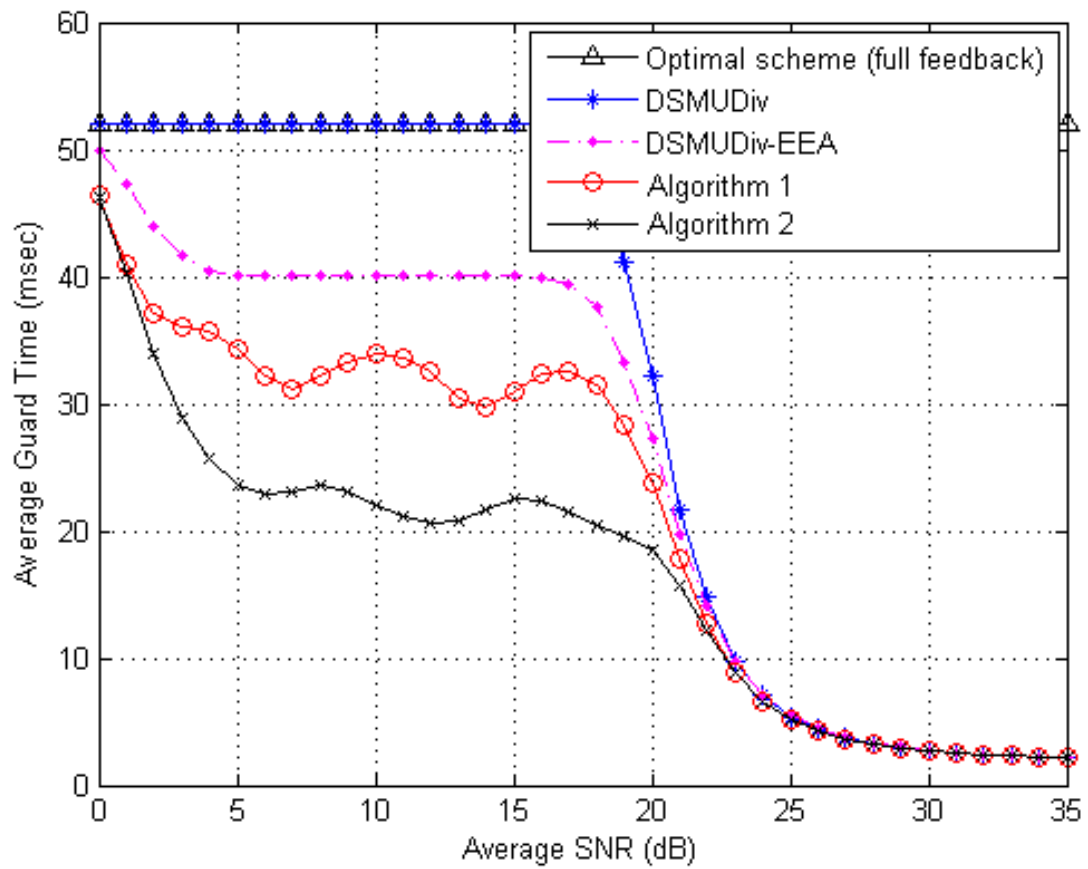


Figure 4.13: Average guard time versus the average SNR. $K = 26$.

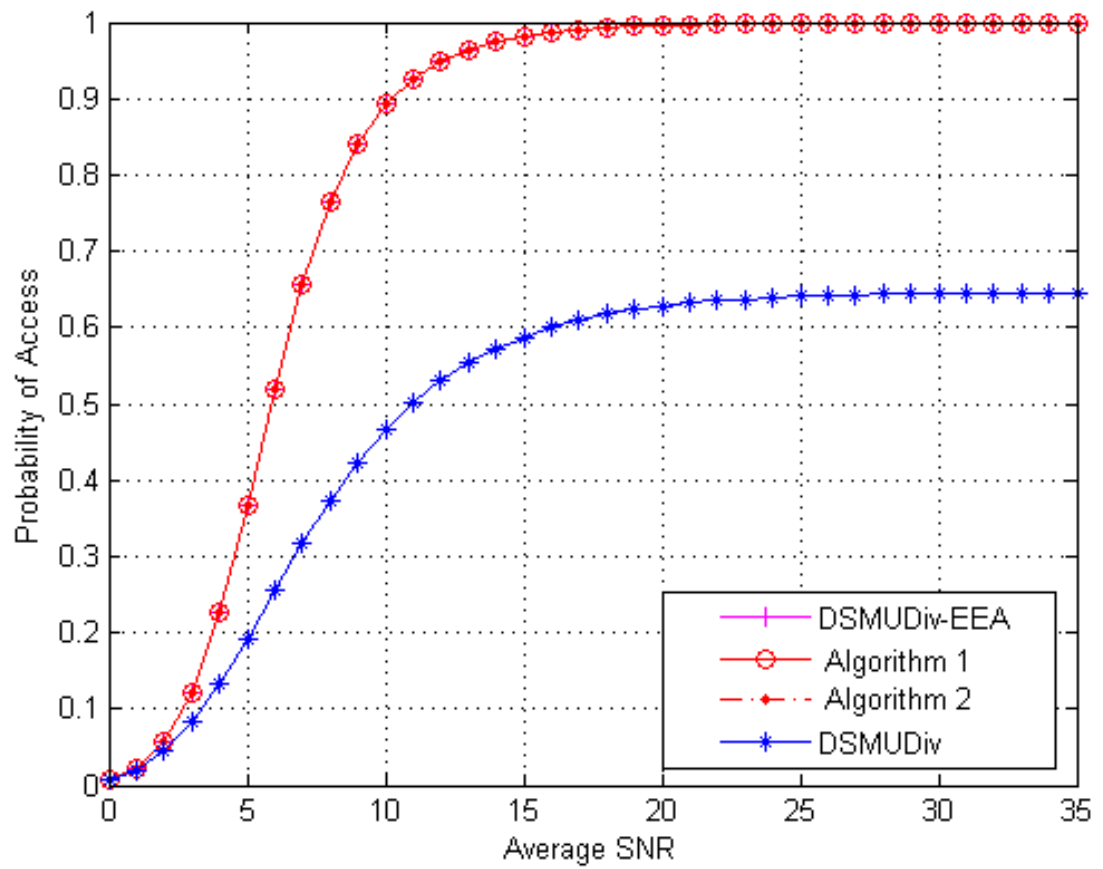


Figure 4.14: Probability of access versus SNR for DSMUDiv scheme, DSMUDiv-EEA scheme and our proposed schemes. $K = 13$.

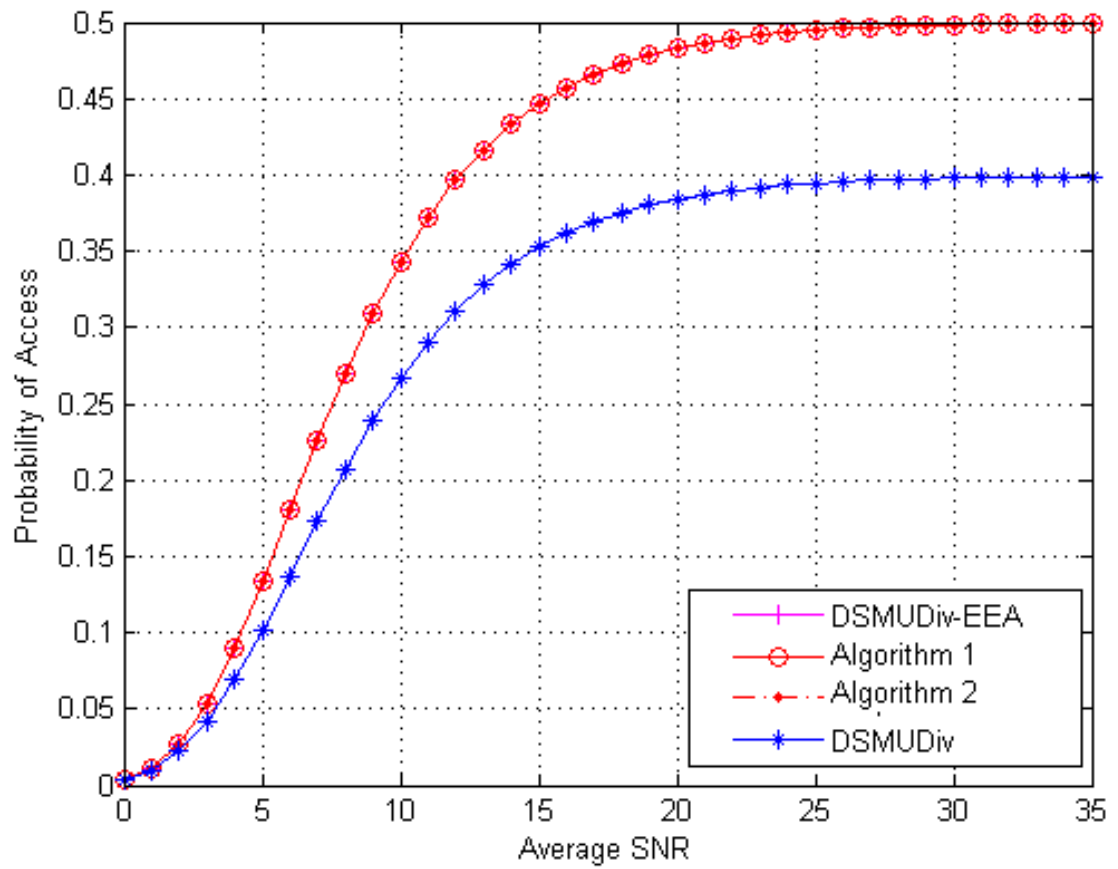


Figure 4.15: Probability of access versus SNR for DSMUDiv scheme, DSMUDiv-EEA scheme and our proposed schemes. $K = 26$.

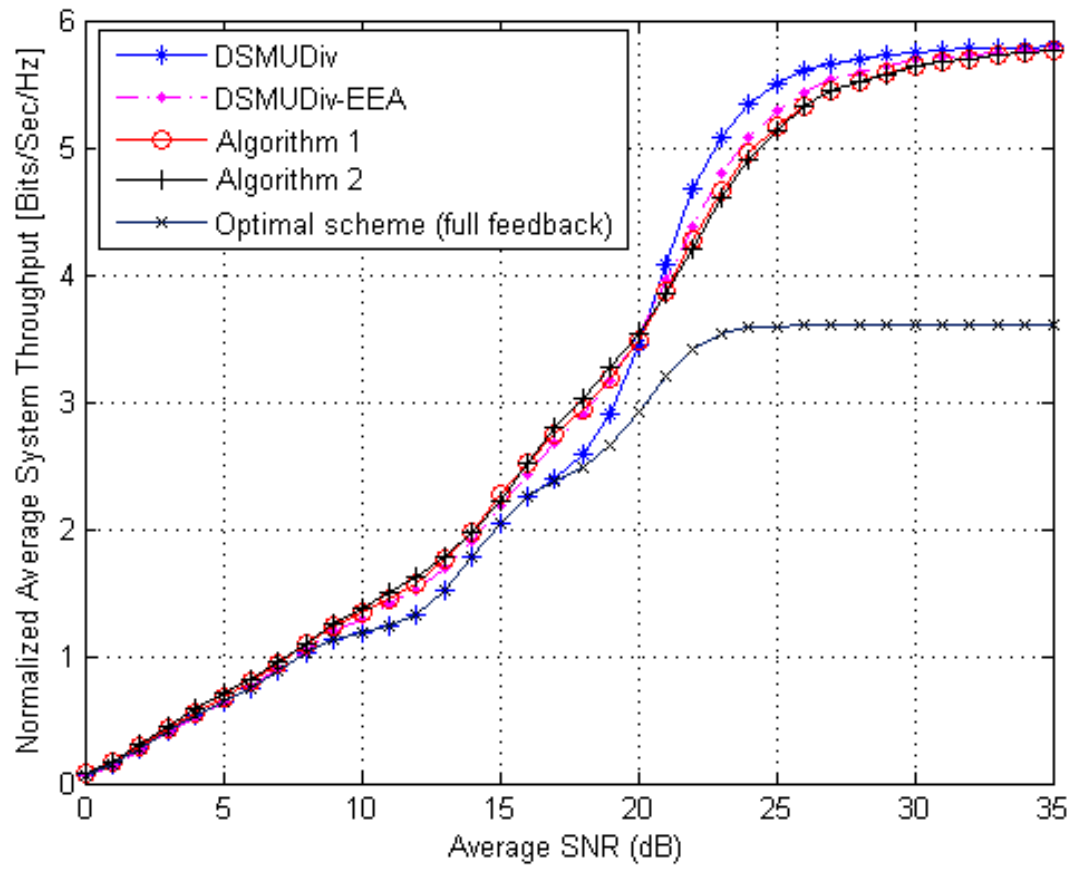


Figure 4.16: Average system throughput for a system with $K = 13$, and $t_d = 5$ msec.

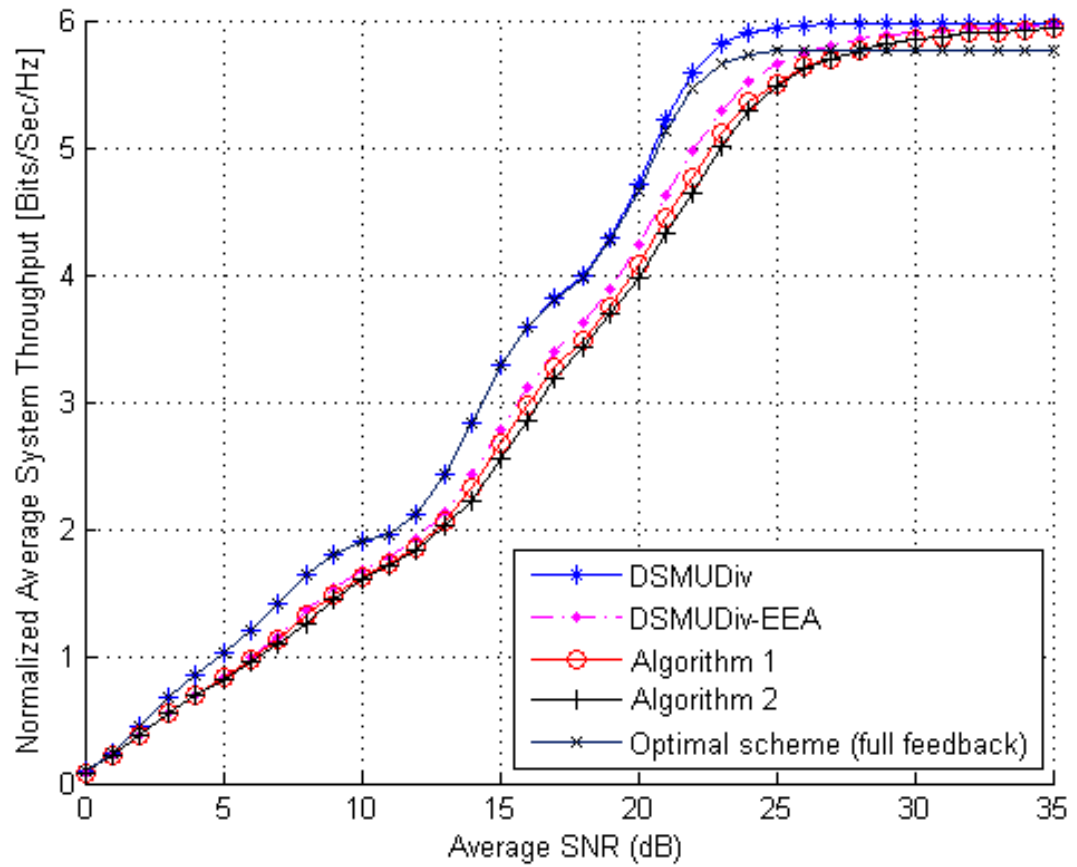


Figure 4.17: Average system throughput for a system with $K = 13$, and $t_d = 50$ msec.

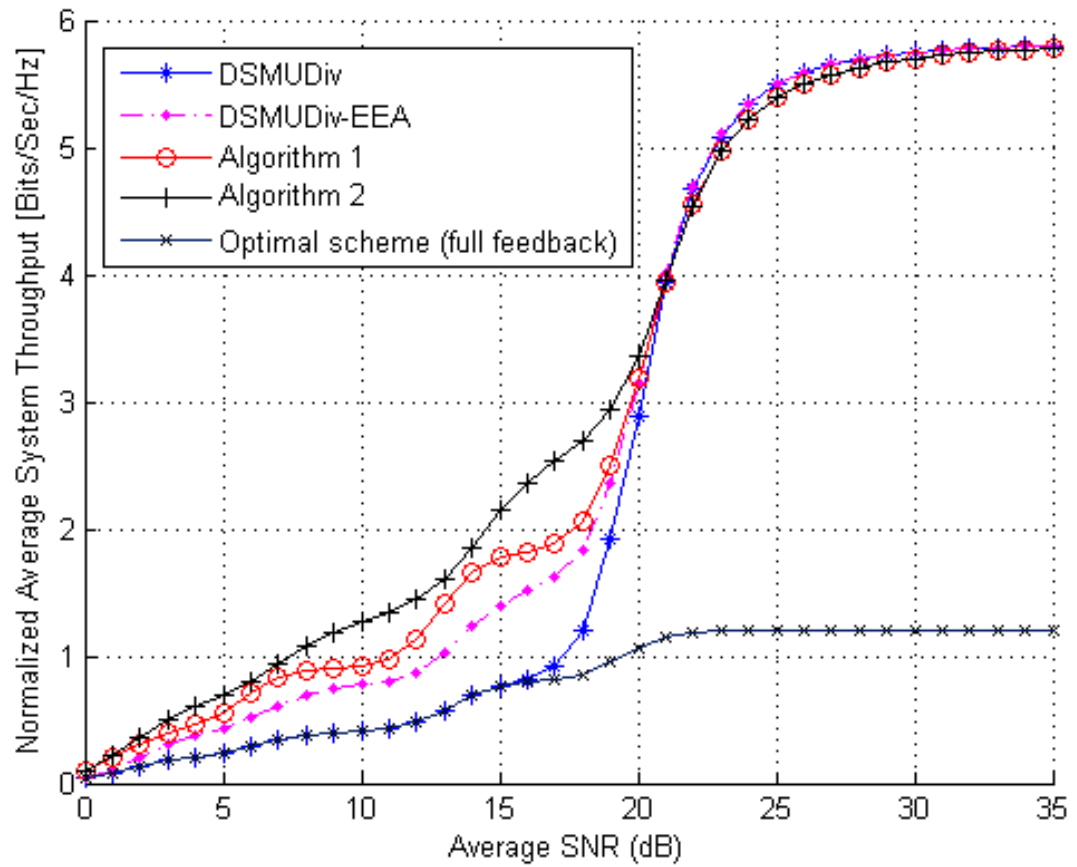


Figure 4.18: Average system throughput for a system with $K = 26$, and $t_d = 5$ msec.

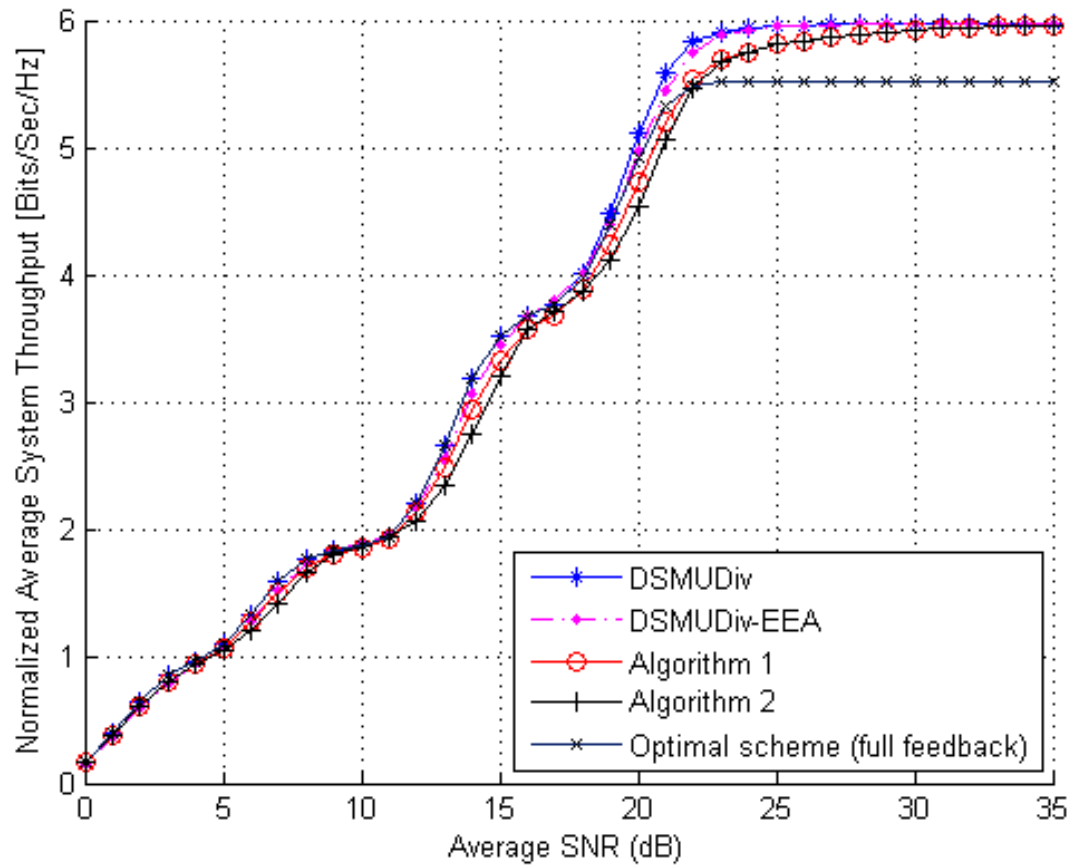


Figure 4.19: Average system throughput for a system with $K = 26$, and $t_d = 50$ msec.

4.6 Conclusion

In this chapter, we proposed two scheduling algorithms that reduce the feedback load in a polling-based multi-carrier system. This was achieved by removing users from the scheduling process once they have been granted a channel resource. Furthermore, we employed multiple adaptive probing thresholds that further reduced the feedback load and scheduling decision time. In addition, we derived closed-form expressions for the average feedback load and the average spectral efficiency. We also studied the effect of scheduling delay on the overall system performance, and we showed that the guard time has a great impact on the average system throughput when the AP transmits over short time-slots. The proposed algorithms have better performance than the Optimal, DSMUDiv and DSMUDiv-EEA schemes when the AP is transmitting data to low and mid average SNR users on short time-slots. Our numerical results showed that the proposed algorithms further reduced the feedback load with slight penalty loss in average spectral efficiency.

Chapter 5

Conclusions and Future Research

The thesis is concluded in this chapter. The thesis summary and the conclusions are presented in Section 5.1. This is followed by suggestions for future research in Section 5.2.

5.1 Thesis Summary and Conclusions

In this thesis, we studied the effect of feedback reduction in a wireless system. We showed that, to optimally schedule users, full feedback was required from all users. However, this resulted in high guard time, scheduling delay and feedback overhead.

In chapter 1, we gave a general background on the wireless channel and some classical diversity techniques. We also presented an extensive literature review on the topic of feedback reduction, and we summarized the thesis contributions.

In chapter 2, we performed a comparative study for some packet scheduling algorithms. We showed that opportunistic algorithms improve the average system spectral efficiency and impose a feedback overhead. We also showed that a tradeoff exists between the system average spectral efficiency and the feedback overhead. Furthermore, we used the Jain fairness index to show that greedy algorithms have poor fairness, i.e. they are not short-term fair. However, in the long run, the fairness index increases.

In chapter 3, we proposed a scheduling algorithm that dramatically reduced the feedback load as compared to the optimal (full feedback) scheme. This was done with no performance degradation. We also showed that only one bit of feedback information was required to schedule two users instead of one user. This one-bit feedback reduced the feedback information payload, and it did not increase much as the number of users increased. Furthermore, we studied the effect of our algorithm on the overall system scheduling delay and throughput. We showed that our algorithm out-performed the optimal scheme and it gave a better performance in both short and long transmission time-slots.

In chapter 4, we introduced two scheduling algorithms that reduced the feedback load in a multi-carrier system with a slight penalty loss in capacity. Despite this slight loss, the feedback reductions produced by our algorithms are much greater than the spectral efficiency loss. Besides this, our algorithms reduced the system average guard time, and thus increased the access probability as compared to the

optimal scheme. This reduces the transmission latency and, as we have shown, it increases the overall system throughput when the system employs short time-slot durations for data transmission.

5.2 Future Research

Future research work can consider feedback reduction schemes for multi-user multi-carrier multi-antenna systems which are gaining much attention recently due to their high data-rates and diversity gains. Additionally, scheduling and feedback reduction schemes for mobile users can be considered, as more feedback information would be required to track the user's channel.

The effect of correlation on a system is known to reduce the overall system performance due to diversity loss. Feedback reduction algorithms can exploit this correlation to reduce the feedback load. Furthermore, research work can seek feedback quantization optimization and determine the possibility of scheduling more users by using only 1-bit feedback information.

Bibliography

- [1] P. Viswanath, D. Tse, and R. Laroia, “Opportunistic beamforming using dumb antennas”, *IEEE Transactions on Information Theory*, vol. 48, pp. 1277- 1294, June 2002.
- [2] T. Rappaport, “*Wireless communications: principles and applications*”, Pearson Education Corporation, Singapore, 2nd edition, 2002.
- [3] D. Tse and P. Viswanath, “*Fundamentals of Wireless Communication*”, Cambridge University Press, September 2004.
- [4] T. Eriksson and T. Ottosson, “Compression of feedback for adaptive transmission and scheduling”, *IEEE Proceedings*, vol. 95, no. 12 pp. 2314-2321, December 2007.
- [5] D. Gesbert and M. Slim-Alouini, “How much feedback is multi-user diversity really worth?”, in *Proc. IEEE International Communication Conference (ICC’04)*, Paris, France, June 2004, pp. 234-238.

- [6] V. Hassel, M.-S Alouini, G. Øien and D. Gesbert, “Rate-Optimal Multi-user Scheduling with Reduced Feedback Load and Analysis of Delay Effects”, *EURASIP Journal on Wireless Communications and Networking*, vol. 2006, no. 26424 pp. 7, 2006.
- [7] V. Hassel, M. Alouni, D. Gesbert, G. Øien, “ Exploiting multi-user diversity using multiple feedback thresholds”, *IEEE Proceedings on Wireless Communications*, vol. 9, no. 5, June 2005.
- [8] Y. Al-Harhi, A. Tewfik, and M. Alouini, “Multiuser Diversity with Quantized Feedback”, *IEEE Transactions on Wireless Communications*, vol. 6, no.1, January 2007.
- [9] S. Jha and M. Hassan, “*Engineering Internet QoS*”, Artech House, 2002.
- [10] B. Sklar, “*Rayleigh Fading Channels in Mobile Digital Communication Systems Part I: Characterization*”, *IEEE Signal Processing Lett.*,*IEEE Communications Magazine*, July 1997.
- [11] B. Sklar, “*Rayleigh Fading Channels in Mobile Digital Communication Systems Part II: Characterization*”, *IEEE Signal Processing Lett.*,*IEEE Communications Magazine*, July 1997.

- [12] R. Knopp and P. A. Humblet, "Information capacity and power control in single-cell multiuser communications", in *Proc. IEEE International Communication Conference (ICC'95)*, Seattle, WA, June 1995.
- [13] S. Catreux, V. Erceg, D. Gesbert and R. Heath, "Adaptive modulation and MIMO coding for broadband wireless datanetworks", *IEEE Communication Magazine*, vol. 40 no.6 pp. 108-115, June 2002.
- [14] F. Florén, O. Edfors, and B.-A. Molin, "The effect of feedback quantization on the throughput of a multiuser diversity scheme", in *Proc. IEEE GLOBECOM*, vol. 1, (San Francisco, CA), pp. 497501, Dec. 2003.
- [15] S. Sanayei and A. Nosratinia, "Exploiting multiuser diversity with 1-bit feedback", in *Proc. IEEE Wireless Communications and Networking Conference (WCNC'05)*, New Orleans, LA, March 2005, pp. 978-983.
- [16] Y. Xue and T. Kaiser, "Exploiting multiuser diversity with imperfect one-bit channel state feedback", in *IEEE Transactions on Vehicular Technology*, vol. 56, no. 1, pp. 183-193, January 2007.
- [17] Y. Al-Harhi, A. Tewfik, and M. S. Alouini, "Multiuser diversity-enhanced equal access with quantized feedback in multicarrier OFDM system", in *Proc. IEEE Vehicular Technology Conference (VTC-Fall'05)*, Dallas, Tx, September 2005, pp. 568-572.

- [18] Y. Xue, T. Kaiser and A. Gershman, "Channel-aware aloha-based OFDM sub-carrier assignment in single-cell wireless communications", *IEEE Transactions on Communications*, vol. 55, no. 5, May 2007, pp. 953-962.
- [19] Y.-J. Choi and S. Bahk, "Selective channel feedback mechanisms for wireless multichannel scheduling", in *Proc. IEEE WoWMoM 2006*, Niagara-Falls, NY, USA, June 26-29, 2006.
- [20] Y.-J. Choi, S. Bahk and M.-S. Alouini, "Switched-based reduced feedback OFDM multi-user opportunistic scheduling", in *Proc. IEEE 16th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC05)*, September 2005.
- [21] Z.-H. Han and Y.-H. Lee, "Opportunistic scheduling with partial channel information in OFDMA/FDD systems", *IEEE Vehicular Technology Conference (VTC)*, vol. 1, pp. 511514, September 2004.
- [22] T. Tang and R. W. Heath Jr., "Opportunistic feedback for downlink multiuser diversity", *IEEE Communications Letters*, vol. 9, no. 10, pp. 948-950, October 2005.
- [23] X. Qin and R. Berry, "Opportunistic splitting algorithms for wireless networks with heterogeneous users", in *Proc. of the 38th Conference on Information Sciences and Systems (CISS '04)*, Princeton, NJ, USA, March 2004.

- [24] H. Koubaa, V. Hassel, and G. E. Øien, “Multiuser diversity gain enhancement by guard time reduction”, in *IEEE Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, New York, NY, USA, June 2005.
- [25] H. Koubaa, V. Hassel, and G. E. Øien, “Contention-less feedback for multi-user diversity scheduling”, in *Proc. IEEE VTC Fall 2005*, Dallas, Texas, USA, September 2005.
- [26] C.-S. Hwang and J. M. Cioffi, “Achieving multi-user diversity gain using user identity feedback”, *IEEE Transactions on Wireless Communications*, vol. 7, no. 8, pp. 2911-2916, August 2008.
- [27] Y. Park, D. Park and D. Love, “On scheduling for multiple-antenna wireless networks using contention-based feedback”, *IEEE Transactions on Communications*, vol. 55, no. 6, pp. 1174-1190, June 2007.
- [28] S. Patil and G. Veciana, “Reducing feedback for opportunistic scheduling in wireless systems”, *IEEE Transactions on Wireless Communications*, vol. 6, no. 12, pp. 4227-4232, December 2007.
- [29] M. Nicolaou, A. Doufexi and S. Armour, “A Selective Cluster Index Scheduling Method in OFDMA”, *IEEE Vehicular Technology Conference (VTC)*, pp. 1-5, September 2008.

- [30] P. Svedman, S. Wilson, L. Cimini Jr. and B. Ottersten, "A simplified opportunistic feedback and scheduling scheme for OFDM", *IEEE Vehicular Technology Conference (VTC)*, vol. 4, pp. 1878 - 1882, May 2004.
- [31] K. Bai and J. Zhang, "Opportunistic multichannel Aloha for clustered OFDM wireless networks", *IEEE Vehicular Technology Conference (VTC)*, vol. 55, no. 3, pp. 848-855, May 2006.
- [32] Y.-J. Choi, J. Kim and S. Bahk, "QoS-aware Selective Feedback and Optimal Channel Allocation in Multiple Shared Channel Enviroments", *IEEE Transactions on Wireless Communications*, vol. 5, no. 11, pp. 3278-3286, November 2006.
- [33] M.-G. Cho, W. Seo and D. Hong, "A joint feedback reduction scheme using delta modulation for dynamic channel allocation in OFDMA systems", *IEEE Transactions on Wireless Communications*, vol. 6, no. 1, pp. 46 - 49, January 2007.
- [34] T.-S. Kang and H.-M. Kim, "Opportunistic Feedback Assisted Scheduling and Resource Allocation in OFDMA Systems", *10th IEEE Singapore International Conference on Communication systems, ICCS 2006*.
- [35] J. Chen, R. Berry and M. Honig, "Limited feedback schemes for downlink OFDMA based on sub-channel groups", *IEEE journal on Selected Areas in Com-*

- munications*, vol. 26, pp. 1451-1461, October 2008.
- [36] T. Eriksson and T. Ottosson, "Compression of Feedback in Adaptive OFDM-Based Systems using Scheduling", *IEEE Communications Letters*, vol. 11, pp. 859-861, November 2007.
- [37] R. Jain, D. Chiu, and W. Hawe, "A quantitative measure of fairness and discrimination for resource allocation in shared computer systems", DEC Research Report TR-301, Digital Equipment Corporation, Maynard, MA, USA, September 1984.
- [38] V. Hassel, M. Alouni, M. Hanssen, G. Øien, "Spectral Efficiency and Fairness for Opportunistic Round Robin Scheduling", in *Proc. IEEE International Conference on Communications, (ICC06)*, Istanbul, Turkey, June 2006.
- [39] J. Holtzman, "Asymptotic analysis of proportional fair algorithm", in *Proc. IEEE Symp. on Personal, Indoor and Mobile Radio Communications (PIMRC)*, vol. 2, San Diego, CA, pp. F-33-F-37, September. 2001.
- [40] M. Corson, R. Laroia, V. Park and G. Tsirtsis, "A new paradigm for IP-based cellular networks", *IT Professional*, pp. 20-29, November/December 2001.
- [41] IEEE Standards Department, "IEEE Std. 802.11. part 11: Wireless Lan medium access control (MAC) and physical layer (PHY) specifications", Technical Report, IEEE, NJ, September 1999.

- [42] V. Hassel, “Design Issues and Performance Analysis for Opportunistic Scheduling Algorithms in Wireless Networks”, PhD. Thesis, Norwegian University of Science and Technology, January 2007.

Vita

- Mohamed Elgaily Eltayeb.
- Received Bachelor of Engineering degree (Honors) in Electronics from Sudan University of Science and Technology in December 2004.
- Worked as a TA at the Electronics Department of Sudan University in March 2005.
- Worked as a Technical Trainer and Software Developer at APTECH Computer Education in May 2005.
- Joined King Fahd University of Petroleum and Minerals as a Research Assistant in September 2006.
- APTECH Certified Trainer.
- Cisco Certified Network Associate.

List of Publications

- M. E. Eltayeb and Y. S. Al-Harhi, “Multiuser Diversity with Binary Feedback”, *submitted to Wireless Personal communications*.
- M. E. Eltayeb and Y. S. Al-Harhi, “Opportunistic Multiuser Scheduling Algorithm For Multi-carrier Wireless Data Systems”, *submitted to IET*.
- T. Y. Al-Naffouri and M. E. Eltayeb, “Opportunistic Beamforming with Precoding for Spatially Correlated Channels”, *to appear in the 11th Canadian Workshop on Information Theory (CWIT 2009)*.