Opportunistic Resource Allocation and Relaying Methods for Quality of Service in the Downlink of Future Cellular Wireless Networks

by

Venkatkumar Venkatasubramanian

Bachelor of Engineering in Electronics and Telecommunications .....University of Mumbai (2002)

Master of Science in Telecommunications ..... Victoria University (2004)

Master of Engineering Research in Telecommunications ..... Victoria University (2006)

Submitted in fulfilment of the requirements for the degree of

DOCTOR OF PHILOSOPHY
(ELECTRICAL ENGINEERING)

at the

Center for Telecommunications and Micro-Electronics,

VICTORIA UNIVERSITY,

Australia.

February 2011

Produced in LaTeX 2_ε

Summary

Wireless communications is on the brink of a major change. New technologies called multiple antenna systems (MIMO) and orthogonal frequency division multiple access (OFDMA) will be put together in the deployment of the next generation of cellular standards known as 4G. Consumers can expect peak data rates up to 160 Mbps. If the user is to have a good network experience with multimedia applications, then consistency in service data rates will be needed.

One of the grey areas in cellular systems is provisioning of these consistent data rates when there are a large number of users in the system. For the upcoming 4G networks, a possible way to address the problem is by judicious user scheduling in OFDMA. This approach to radio resource management forms the main theme of the research that is presented in the thesis. To effectively characterise the problem we identify a key requirement of maximising the number of users with guaranteed data rates.

While consistency of data rates is an issue, our tests in OFDM cells also confirm that in certain locations coverage can be well below International Mobile Telecommunications (IMT)-Advanced targets. A particular case of interest is outdoor-to-indoor coverage, where building penetration loss can negatively impact the received signal power. To extend coverage we consider the deployment of relays. Importantly, we show that to obtain benefit from relays resource management functionality is essential at the relays.

The main contributions are the proposal of an OFDMA resource allocation scheme called Load Balanced Opportunistic Resource Allocation (LBORA) and a relaying scheme called Guaranteed Bit Rate by Relay Scheduling (GBRS). The schemes are designed to maximise the number of users with guaranteed data rates. LBORA is applicable to the direct downlink of macro-cell whereas GBRS extends LBORA by including resource allocation at relays.

To obtain realistic results, we collect channel measurements from a LTE-like base station using a transceiver prototype at 2.6 GHz frequency. The link supports two spatial streams (2×2 MIMO) of OFDM. The test-bed was available at Heinrich Hertz Institute in Berlin.

Both non-guaranteed and guaranteed bit rate allocations are described in Chapter 4. For guaranteed bit rates the problem of maximising the number of users is formulated. A linear program is shown to solve the problem. Yet for real-time deployment, faster solutions which are also easy to implement are preferable. With this in mind we present a heuristic scheme called LBORA. The key idea is to integrate admission control with resource allocation, called joint admission control and resource allocation. Numerical results based on simulations from measured channels showed a gain of 78% in the number of users for data rates as compared to a disjoint approach where admission control and resource allocation are done sequentially. A running numerical example is provided throughout the chapter for the sake of illustration.

A relay deployment concept using decode-and-forward relaying is proposed in Chapter 5. The relay has two units : a feeder unit normally mounted outside a building and an access unit normally mounted inside the building. The relay feeder and access units co-share the bandwidth by division of time slots. The idea here is to enable full bandwidth reuse on the access link of relay. Real-time measurements compare coverage with and without a relay when the serving base station is outdoors. The measurements were conducted in indoor office building at Heinrich-Hertz Institute. Results show that coverage holes do exist in macro-cells. In contrast, indoor relays provide excellent coverage when frequency dependent link adaptation is applied.

In Chapter 6, we look at combining resource management at relays and the base station. The relaying scheme has five key functionalities : enabling full bandwidth indoors, duplex time sharing, frequency dependent link adaptation, multiuser scheduling and activation of many relays. These building blocks are put together to create a novel scheme called GBRS, which caters for guaranteed bit rates to users. Hierarchical OFDMA is employed to support multiple relays and multiple users per relay in dedicated time slots. An optimisation framework is proposed for maximising the number of users. Optimal and heuristic solutions to GBRS are provided. Numerical results are shown in various indoor test scenarios for a motivational 2 Mbps data rate for each user. Best case results show that with 10 active relays 40 users can be supported whereas without any relay direct downlink can only support 20 users.

A clear strategy to support the remaining 20 users in the absence of relays would be doubling of the spectrum. Spectrum however comes at a price which might happen to be substantially more than the relays. Thus we expect relays to be very beneficial for the roll out of future OFDMA cellular networks.

Declaration

I, Venkatkumar Venkatasubramanian, declare that the PhD thesis entitled "Opportunistic Resource Allocation and Relaying Methods for Quality of Service in the Downlink of Future Cellular Wireless Networks" is no more than 100,000 words in length including quotes and exclusive of tables, figures, appendices, bibliography, references and footnotes. This thesis contains no material that has been submitted previously, in whole or in part, for the award of any other academic degree or diploma. Except where otherwise indicated, this thesis is my own work.

Signature Date

Acknowledgements

First of all, I would sincerely like to express my gratitude to my supervisor, Prof. Mike Faulkner for his continual support. I have benefitted from his inspiration and dedication to the growth of his students. Many thanks. I would also like to thank Dr. Thomas Haustein at Heinrich Hertz Institute in Berlin for motivation and countless efforts. I want to thank him for his interactions and sharing his broad perspective to this research. Thanks a ton.

I also wish to give my special thanks to all my colleagues, at both Victoria University in Melbourne and at Heinrich Hertz Institute in Berlin. Cheers to Andreas Forck, Thomas Wirth and Holger Gaebler for intense dedication to the measurement campaigns and the test-bed. It is in fact very inspiring to see such engineering efforts. Thanks also go to other colleagues in the test-bed team.

Special thanks to our secretary Susanne Muller for providing the logistical support at Berlin and during my travels. Thanks also to all my other friends in Berlin for making it a truly wonderful time.

A special word of thanks to Dr. Egon Schulz and the other colleagues at Nokia Siemens Networks in Munich during the internship. It was indeed a time well spent in that magnificient city.

Of the many friends and colleagues in Melbourne, I would also thank Dr. Himal Suraweera and Dr. Terence Betlehem for the talks. Special regards to Dr. Aaron Reid and Matthew Williamson for the unforgettable times, several well-timed discussions and the cricket matches.

Most of all, I am especially grateful to my parents and sister for their tremendous love to help me complete this thesis.

Related Publications

Peer Reviewed

1. V. Venkatkumar, T. Wirth, T. Haustein and E. Schulz, "Relaying in Long Term Evolution: Indoor full frequency reuse," Proceedings of IEEE European Wireless Conference 2009, pp. 298-302, Aalborg, 17-20 May 2009.

2. V. Venkatkumar, T. Haustein and M. Faulkner, "Relaying results for indoor coverage in long-term evolution and beyond," European Transactions on Telecommunications, Special Issue on European Wireless 2009 , vol. 21, issue 8, pp. 770-779, December 2010.

3. V. Venkatkumar, T. Haustein and M. Faulkner, "Joint Admission Control and Resource Allocation for Multiuser Loading in LTE networks," *Smart Spaces and Next Generation Wired/Wireless Networking*, Lecture Notes in Computer Science, Volume 6294/2010, pp. 421-435, 2010.

4. V. Venkatkumar and T. Haustein, "Multi-user Relaying with Full Frequency Reuse for Enhanced LTE-GBR Coverage," *Proceedings of European Wireless Conference 2011*, Vienna, Austria, Accepted for publication.

5. V. Venkatkumar, T. Haustein et. al, "Field Trial Results on Multi-User MIMO Downlink OFDMA in Typical Outdoor Scenario Using Proportional Fair Scheduling," International ITG Workshop on Smart Antennas, 2008. WSA 2008, pp. 55-59, Darmstadt, 26-27 February 2008.

6. T. Haustein, V. Venkatkumar et. al, "Measurements of Multi-Antenna Gains using a 3GPP-

LTE Air Interface in Typical Indoor and Outdoor Scenarios," *Proceedings of European Wireless 2008*, pp. 1-6, Prague, 22-25 June 2008.

7. T. Wirth, V. Venkatkumar, T. Haustein, E. Schulz and R. Halfmann, "LTE-Advanced Relaying for Outdoor Range Extension," Proceedings of IEEE VTC 2009-Fall, pp. 1-4, Anchorage, USA, 20-23 September 2009.

8. T. Wirth, V. Venkatkumar et. al, "Polarisation Dependent MIMO Gains on Multiuser Downlink OFDMA with a 3GPP LTE Air Interface in Typical Urban Outdoor Scenarios," International ITG Workshop on Smart Antennas, 2008. WSA 2008, pp. 157-161, Darmstadt, 26-27 February 2008.

Other works

1. V. Venkatkumar and M. Faulkner, "Energy-efficiency of Multiple-Relay Cooperation in Sensor Networks," RUNES Summer School, London, UK, 9-11 July 2007. http://www.ist-runes.org

2. A. Forck, T. Haustein, V. Jungnickel, V. Venkatkumar, S. Wahls, T. Wirth, and E. Schulz, "Early real-time experiments and field-trial measurements with 3gpp-lte air interface implemented on reconfigurable hardware platform," 3GPP LTE Handbook: 3GPP LTE Radio and Cellular Technology, Editors B. Furht and S. A. Ahson ch. 11, pp. 365411, CRC Press, Auerbach Publications Boston, MA, USA, 2009.

Contents

List of Figures

List of acronyms

3GPP	third generation partnership project
4G	fourth generation
AF	amplify and forward
AWGN	additive white gaussian noise
BC	broadcast channel
BER	bit error rate
CDMA	code division multiple access
CDF	cumulative distribution function
DF	decode and forward
EPC	evolved packet core
ETSI	european telecommunications satandards institute
E-UTRAN	evolved universal terrestrial radio access network
FDD	frequency division duplexing
FDMA	frequency division multiple access
FFT	fast fourier transform
GBR	guaranteed bit rate
GBRS	guaranteed bit rate by relay scheduling
HARQ	hybrid automoatic repeat request
ITU	international telecommunications union
LTE	long term evolution
LBORA	load balanced opportunistic resource allocation

MAC	medium access control
MIMO	multiple input multiple output
MRC	maximum ratio combining
MCS	modulation and coding scheme
OFDM	orthogonal frequency division multiplexing
OFDMA	orthogonal frequency division multiple access
PF	proportional fairness
PSK	phase shift keying
PER	packet error rate
QPSK	quadrature phase shift keying
QAM	quadrature amplitude modulation
SNR	signal to noise ratio
TDMA	time division multiple access
TCP	transmission control protocol
TTI	transmit time interval
V-BLAST	vertical bell laboratories layered space-time architecture

List of symbols

x_{km}	Resource allocation variable of resource block m to user k. Takes values between 0 and 1.
u_{km}	Spectral efficiency of resource block m for user k in terms of bits per subcarrier.
Δ_r	Variable number of resource blocks used for sending data to relay r.
Ω_k	Data rate demand of user k in a scheduling interval
K	Total number of users
M	Total number of resource blocks
L	Number of subcarriers in a resource block
T	Number of OFDM symbols in a transmit time slot
N	Total number of time slots used in a relaying radio-frame
r_k	Data rate allocated to a user k after scheduling
a_k	Binary variable which indicates if a user has been provided the data rate
γ_r	Spectral efficiency of the feeder link to relay r
R	Total number of active relay cells within a macrocell and in a scheduling interval.
ϵ_r	Normalised number of time slots used by the access link of relay r.
K_r	Active set of users at relay r
F_r	Feasible set of users at relay r
e_{kv}	Binary variable which indicates if the feedback level v is allocated to user k.
L_k	Bandwidth metric showing the estimated amount of bandwidth needed by user k.

Chapter 1

Introduction

1.1 Future cellular communications

Wireless communications is a branch of telecommunications which concerns information transfer without using wires. Today's notion of anytime, anywhere communication may not have been possible without wireless communications. Wireless communication comprises a broad and different number of technologies and standards. One of them is cellular technology which aims at reliable connectivity over a wide area using a licensed spectrum. The data rate capability of cellular communications has grown tremendously, from a few kilo bits per second in 1970 to few megabits in 2010. Surprisingly, even the current capabilities have proven not to be sufficient for user requirements. The next generation of networks is about to be rolled out.

The migration from the current 3G to the next generation 4G is expected to happen over the next few years. The ultimate goal is to provide ubiquitous and high quality wireless access in the form of high data rate, low latency transmission for application such as voice over IP, and uninterrupted video streaming for users. On a conceptual basis, 4G systems envisage to seamlessly integrate old and new terminals using multiple cellular standards [1],[2]. For a cellular operator, the main challenge will be to successfully deliver these services to the users through proper deployment methods. We now describe some recent advances.

1.2 Overview of the enabling technologies

Cellular technology has made many changes over the years in order to meet the requirements of users and applications. At the time of the roll out of second generation (2G) networks, the main application was voice telephony. The backbone network and the radio access network of 2G networks (e.g, Global System for Mobile Communications (GSM)) were tailored to the requirements of voice traffic. For third generation (3G) networks, the attention was more towards the Internet data traffic, in the form of web pages and file downloads. Therefore, larger bandwidths and higher data rates were the additional motivations.

For beyond 3G networks, new types of applications are on the rise. Specifically, real-time video streams and on demand streaming, such as youtube videos are now available for users. The demands placed by these applications on the wireless network are substantial. In most scenarios, a mixture of different traffic classes may have to be catered for e.g, a user downloading a file and watching a video at the same time. These types of traffic situations demand flexibility from the wireless network. To provide flexibility, a few fundamental features of the upcoming networks have been reworked. The most notable developments have been the following :

- a physical layer based on orthogonal frequency division multiplexing (OFDM);
- a low latency backbone architecture, a low round-trip time radio access layer and Internet protocol (IP) based ;
- a more user-centric approach, wherein end-user experience is an objective and supported using receiver to transmitter feedback ;

Motivated by these possibilities, a new 3^{rd} generation partnership program (3GPP) [3] standard called long term evolution (LTE) [4] has been developed as pre-4G. 3GPP is a collaboration agreement, established in December 1998, that brings together a number of telecommunications standards bodies, known as Organizational Partners. The current Organizational Partners are Association of Radio Industries and Business (ARIB), China communication standards association (CCSA), European Telecommunications Standards Institute (ETSI), Alliance for Telecommunications Industry Solutions (ATIS), Telecommunications technology association (TTA) and The Telecommunications technology committee (TTC). Researchers and development engineers from all over the world representing more than 60 operators, vendors and research institutes are participating in the joint LTE radio access standardization effort.

The starting point for LTE standardization was the 3GPP radio access network evolution workshop, held in November 2004 in Toronto, Canada.

The proposal is to use an orthogonal frequency division multiplexing (OFDM) air interface together with advanced antenna technologies. Moreover LTE supports flexible carrier bandwidths, from below 5MHz up to 20MHz and both FDD (Frequency Division Duplex) and TDD (Time Division Duplex). The peak downlink data rate has been increased up to 160 Mbps in 20 MHz bandwidth.

The main benefit from OFDM is that the communication bandwidth between the transmitter and receiver is decomposed into parallel channels, each of narrow bandwidth. The intention is that the signal can be retrieved with relative ease at the receiver. This possibility also helps to apply advanced antenna concepts to a large bandwidth. In order to facilitate this working, an OFDM transmitter inserts guard bands known as cyclic prefix between the time symbols. Through this mechanism of cyclicity in the signal, equalisation is performed individually for each narrowband channel.

In the network architecture, the role of a base station has become more important. In effect each base station handles resource allocation and retransmissions by itself instead of relying on a centralised radio resource controller. Thus the term flat architecture. Each base station communicates directly to the network gateway, which is a component of the so called evolved packet core network (EPC). This architecture is defined as part of the System Architecture Evolution (SAE) effort.

Overall, a study item proposed in December 2004 [5] decided the following important objectives for realising the success of LTE [6]

- significantly increased peak data rate (100 Mbps), increase the cell edge bit rate whilst using the same base station sites ;

- flexible spectrum usage by scalability from 1.25 MHz to 20 MHz ;
- efficient support of the various types of services ;
- reasonable terminal power consumption ;
- radio-access network latency below 10 ms ;
- system should be optimized for low mobile speed but also support high mobile speed.

The study item further kept track that the LTE concept could fulfill a number of requirements specified in 3GPP TR 25.913 [6] through feasibility study of evolved universal terrestrial radio access network (E-UTRAN). A key protocol in the EPC is Internet protocol (IP) which enables integration of all kinds of applications to the nodes in the backhaul and radio access network. The IP functionality allows the integration of different traffic classes such as voice and data traffic from the Internet as IP packets. LTE performance has been evaluated in so called checkpoints and the results were agreed on in 3GPP plenary sessions during May and June 2007 in South Korea [7]. The results show that LTE meets, and in some cases exceeds, the targets for peak data rates, cell edge user throughput and spectrum efficiency.

1.3 Challenges

In spite of these promising initial results, there are numerous aspects that need investigation for successful deployment. Two important aspects are quality of service provisioning and coverage. In the framework of LTE, one of the main components in the network architecture is multi-user scheduling. This functionality can be used to design efficient solutions to achieve the basic targets : flexible spectrum usage, reduced cost per bit and high spectral efficiency.

To achieve the goal of high quality of service, we investigate suitable scheduling algorithms and concepts. We identify a particular issue in this thesis : how to maximise the number of users supported at a specified data rate. As we show later, there are numerous intricacies that needs to be taken into account in designing a scheduling algorithm.

LTE trial results have shown the realisation of a high data rate in both peak and average measures [7], [116]. However, recently completed simulation studies have revealed possible coverage related problems at outdoor cell edge and indoor environments [11]. Therefore, an experimental verification will be done to establish if these issues exist in real-time environments. Specifically, this thesis considers indoor cells.

Meanwhile International Telecommunications Union (ITU) has looked even further and set high spectral efficiency targets for the upgrade of 4G systems such as LTE-Advanced [12]. An average cell spectral efficiency requirement of 3 bits/s/Hz/cell for indoor users has been set for fourth generation systems (4G) in [13]. A user spectral efficiency target of 0.1 bits/s/Hz is also recommended by [13]. These targets can indeed be difficult to meet when there are many users and especially if coverage issues exist in a cell.

One possible way to solve coverage issues is by using relays which were widely investigated for enhanced 3G called high speed downlink packet access (HSDPA) networks. These networks use code division multiple access (CDMA). Because CDMA and OFDM are different interfaces, the investigation of relaying concepts and deployment in OFDM networks requires a new insight. One prerequisite is that relaying need not only be beneficial but also flexible, in the sense that it can be easily integrated into the system architecture. In this thesis, we look at the above-mentioned issues of multi user scheduling and indoor relaying for future cellular systems.

1.4 Contribution

The main contributions of this thesis include :

- **MIMO-OFDM measurement results**. We conduct real-time measurements on a MIMO-OFDM downlink interface with the help of a test-bed. The results show that 3GPP target cell spectral efficiency can be achieved using current hardware capabilities. Simulation studies further support the intuition that average spectral efficiency can be improved by using multi-user scheduling.

- **Scheduling framework**. A new framework for quality of service provisioning in the downlink of LTE is presented in Section 4.3. Specifically, a performance metric known as 'maximising the number of users' is proposed for supporting guaranteed bit rate services which is a traffic class in LTE. The scheduling framework is beneficial in the spirit of a user centric approach to a cellular system. A new concept called 'joint admission control and resource block allocation' is introduced to support the scheduling framework.

- **Proposal of new scheduling techniques**. Optimal and suboptimal methods are proposed for the scheduling framework when operating in the guaranteed bit rate mode. The solutions are derived analytically in Sections 4.3.3 and 4.3.4 respectively. The relation to linear programming and assignment problems is revealed. A novel heuristic scheduling scheme called Load Balanced Opportunistic Resource Allocation (LBORA) is presented in Section 4.3.4.4. Numerical illustrations of the schemes are provided throughout the chapter for clarity.

- **A relaying deployment concept for OFDM**. A relaying deployment concept is proposed for 4G systems to solve coverage issues which may arise in MIMO-OFDM for carrier frequencies above 1 GHz. It is notable that a framework of relaying is expected to be a key component in LTE-Advanced networks and some standardisation work is already underway [15]. The relay architecture we propose is equipped with multiple transmit and receive antennas and is of decode and forward type. Analytic characterisation of the achievable throughput with the relay is shown via optimisation of the duty cycle (active time) of the relay.

- **Measurement studies with and without relaying**. We further conduct real time measurements for the proposed relaying concept. For this purpose, we use a relay prototype which can handle real time streaming transmissions. The basic features of a relaying prototype which implements our decode-and-forward relay concept is shown. Coverage studies are conducted with and without relaying in an indoor environment. The results show achievable throughput with link adaptation. We conclude from the measurements that a) coverage holes do exist in indoor environment and b) multi antenna relaying is beneficial for solving the problem. These real-time results are the first of their kind.

- **Simulation based results from real-time channels**. Rigorous results for quality of service metrics with LTE parameters are shown using simulations. The simulations are perfomed using offline analysis of the real time channel values which are captured during the measurements. Performance of the scheduling and relaying schemes are simulated from the stored channels. A classification of three feedback schemes for link adaptation is made on the basis of their overheads. Logical and intuitive explanations for the performance are provided with respect to feedback, and is observed to firmly support the simulation studies. These intuitions are further utilised in developing a new feedback framework called 'User Selective Feedback' when the uplink capacity is limited.

- **Multi-user relaying scheme**. A new relaying scheme called 'Guaranteed Bit Rate by Relay Scheduling' (GBRS) is presented to improve quality of service and coverage in a multi-user multi-relay scenario. The scheme exploits resource allocations to users at each relay and duty cycle optimisation of each relay. A second resource allocation is employed at the base station. A novel method for optimisation of the GBRS scheme is presented, which employs two steps : relay resource allocation (RRA) and group selection (GS). The relation of optimal RRA solution to a fractional linear program is revealed. It is shown that the working of the GBRS scheme can be realised as a hierarchical admission control at the relays and the base station.

- **Simulation test results for multi-user relaying**. Numerical results are shown for the multi-user multi-relay scenario using offline analysis of the stored set of channels from the measurements. Four test scenarios are conceived on the basis of varying number of users, varying number of relays and for different link packet error rates. Simulation results are shown for the test scenarios, along with logical explanations which firmly support the results.

1.5 Structure of the thesis

The thesis is organised as follows.

Chapter 2 provides a first introduction about Orthogonal frequency division multiple access (OFDMA) and multiple antenna concepts of 4G-LTE systems. The OFDMA experimental test-bed which has been used for collecting channel measurements is described. The key challenges that were met in the implementation of the test-bed are described. The test-bed has been used for the purpose of verification of the concepts in the thesis.

Chapter 3 presents a literature survey on OFDMA resource allocation and relaying. Some of the existing approaches for providing quality of service and their potential shortcomings are highlighted.

Chapter 4 concerns the scheduling methods than can be employed for non-guaranteed and guaranteed bit rate allocation in OFDMA. Heuristic schemes are presented for rate proportional and utility proportional fair objectives in the case of non-guaranteed bit rates. The system framework of guaranteed bit rates is introduced, which can employ a 'joint admission control and resource block allocation' mechanism. Optimal and suboptimal scheduling schemes are presented for the proposed mechanism. Simulation results of the proposed scheduling schemes are demonstrated from measured channels. The suboptimal scheme named 'Load Balanced Opportunistic Resource Allocation' is shown to achieve close-to-optimal performance with low complexity.

Indoor relay measurements using a decode-and-forward relay is reported in Chapter 5. The real-time results of relaying is compared to that of outdoor to indoor coverage from a base station without a relay. A classification of feedback on the basis of overhead is drawn in this chapter. Simulation based results are shown which better reveal the impact of feedback on coverage, with and without a relay. Finally, a novel feedback scheme called user selective feedback is presented in the view of its benefit to deployment.

Chapter 6 proposes a new relaying scheme called 'Guaranteed Bit Rate by Relay Scheduling' for supporting guaranteed bit rates in a relay scenario. The problem formulation is shown for multi-user multi-relay scenario. Optimal and sub-optimal solutions are given using the functional steps of relay resource allocation and grouped admission of relays and users. Finally, numerical results of the proposed concepts are shown which confirm that the proposed relaying scheme can realise significant gain in indoor cells.

Chapter 2

4G features

Resource allocation in cellular

Resource allocation in wireless networks distributes the units of time, bandwidth and power to different communication links. In cellular networks, the communication links are mainly the radio interface between base stations and user equipment. Transfer of information from a base station to user equipments is said to be the downlink direction and the reverse direction is called the uplink.

In a cellular downlink, the base station acts as the sender and the user equipment is the receiver. Typically, in a given geographic location, an amount of bandwidth is licensed by an operator. For the purpose of management, the licensed bandwidth may be subdivided and assigned to different base stations based on a frequency reuse factor. This process of resource management is referred to as cell planning or network planning. Many early studies have advocated the usage of frequency reuse factor 1 which means that all the base stations transmit on the same frequency. It is notable that baseline system level simulation assumption for LTE performance assessment in [7] is a reuse factor of 1. However, there is no tight requirement as such on the type of frequency reuse scheme and thus is a design choice for radio resource management. For example, LTE performance assessments comparing different fixed frequency reuse factors have been separately conducted in [8], [9], wherein the results of [8] favour reuse 1 and [9] favours a fixed reuse factor depending on cell size to realise gains in average and cell edge throughputs. A mixture of reuse factors has been independently investigated, for example the fractional frequency reuse [17] scheme, which in effect employs a reuse factor for cell edge users.

In the evolution of LTE towards LTE-Advanced (4G), it is expected that additional spectrum would be available for the operators. Therefore, it is quite possible that 4G networks would implement some amount of frequency planning, in order to reduce interference from adjacent macro-cells. The measurements we later report in the thesis are conducted in such an interference-free scenario.

In each base station, a second level of radio resource management is employed to serve multiple users in the downlink. We refer to this second level of resource management simply as 'resource allocation'. In the upcoming radio interface, by using multiple antennas at the base station and the user equipment, the so called multiple input multiple output system (MIMO) can be realised. In what follows, we introduce two main features of 4G networks : OFDMA and MIMO and later describe the MIMO-OFDMA test-bed.

2.1 OFDMA

OFDMA stands for orthogonal frequency division multiple access. The principle of OFDMA is to partition the subcarriers to different users within an OFDM signal. For resource management, a set of adjacent subcarriers in a time slot of 0.5 ms is defined as a resource block. The possibility in OFDMA is to schedule a different user on a resource block every 1 ms. This time duration of 1 ms is referred to as transmit time interval (TTI) or a subframe. Thus different users can share the entire bandwidth in a given transmit time interval (which consists of two 0.5 ms time slots) as shown in Figure 2.1. A block of 10 TTIs is called a radioframe. To enhance downlink performance, OFDMA can use feedback in the uplink which conveys information regarding a user's downlink channel quality. By using this feedback for resource management, the downlink can realise the so-called multiuser diversity [44]. Multiuser diversity is a selection diversity technique that selects a user to be served on a resource block on the basis of channel quality. This intuitively makes use of the independence of the channel of different users because the users are spatially separated. Simply speaking, the channel (resource block) can be allocated to the user who experiences the best

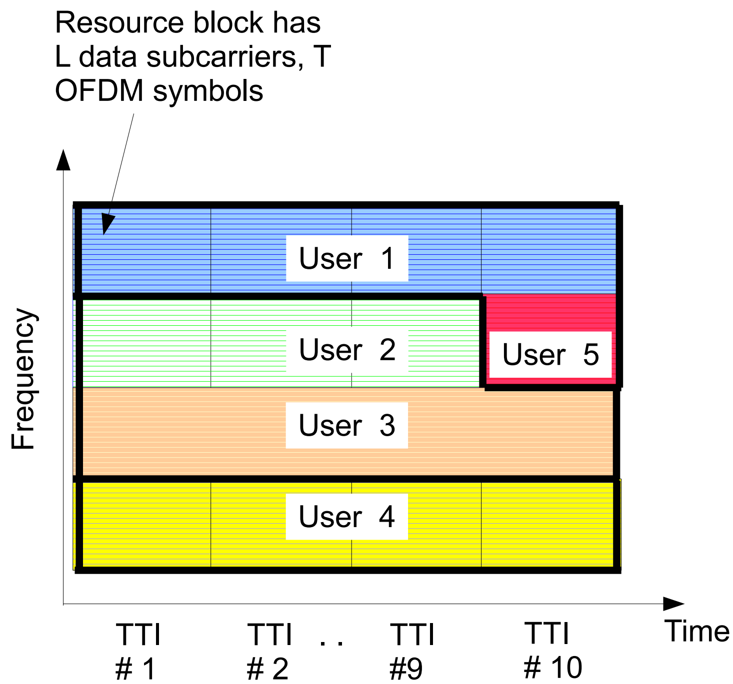


Figure 2.1: Illustration of resource allocation in OFDMA. Resource blocks and transmit time intervals (TTIs) are allocated to different users.

channel conditions. A particularly attractive feature is that this type of opportunistic allocation can be done in OFDMA on each resource block separately. A drawback of a totally greedy allocation however is that all the subcarriers may get allocated to one user who has the best channel to the base station. This can result in data rate starvation for all other users. To address this issue, fairness is needed. A simple way of providing fairness is a round robin fairness, which is often called the round robin scheme. More sophisticated understanding of fairness and some methods are discussed in the literature study in Section 3.1.

It is worthwhile to note that a standout feature in 3G and beyond 3G systems is adaptive modulation. A significant novelty of LTE and 4G is that adaptive modulation is coupled to the idea of user-resource block mapping. Thus it becomes possible to assign a resource block to a user such that a higher modulation order can be transmitted on the subcarriers of that resource block.

Two other resources of time and power can be utilised by : a) serving a different set of users in each time slot and b) performing power allocation in a resource block and user dependent fashion.

Heteregenous network elements

Heterogenous networks elements in the form of remote radio heads, picocells, microcells and femtocells along with macro base stations are expected in the 4G networks. These are mainly hotspot deployments. According to 3GPP [15], "Heterogeneous deployments consist of deployments where low power nodes are placed throughout a macro-cell layout. For such depolyments, picocell base stations are mainly installed to provide outdoor hotspot coverage, whereas femto-cells base stations provide in building coverage. The current view is that these low power nodes will be connected to a wireline backhaul network. 3GPP is working on enhanced inter-cell interference coordination mechanisms in time domain and frequency domain for carrier aggregation based and non-carrier aggregation based schemes [15] to handle heterogenous nodes.

According to 3GPP [15], relaying is considered for LTE-Advanced as a tool to improve e.g. the coverage of high data rates, group mobility, temporary network deployment, the cell-edge throughput and/or to provide coverage in new areas. One main difference of relays with respect to other heterogenous elements is that the relay node is wirelessly connected to the radio-access network via a donor cell. The relays wirelessly connect to the macrocell also for all signalling and for this reason they may be viewed upon as radio resources belonging to a macro-base station. The scope for 'relays' within 3GPP for 4G networks specifically mentions regenerative relays. Regenerative relays will be future alternatives to conventional booster-type relays called repeaters.

Put in a nutshell resource allocation in 4G OFDMA can use all of the following: power allocation, time slot scheduling, subcarrier allocation and adaptive modulation to a user along with relays. In Figure 2.2 we present an evolution of cellular technologies with respect to downlink resource allocation.

A list of 3GPP performance targets for the radio interface in low mobility scenario of 3 Kmph as obtained from [6],[7],[12] is presented in the Tables 2.1 and 2.2.

ITU has independently set the following requirements for IMT-Advanced (4G) for different environments as in Table 2.2.

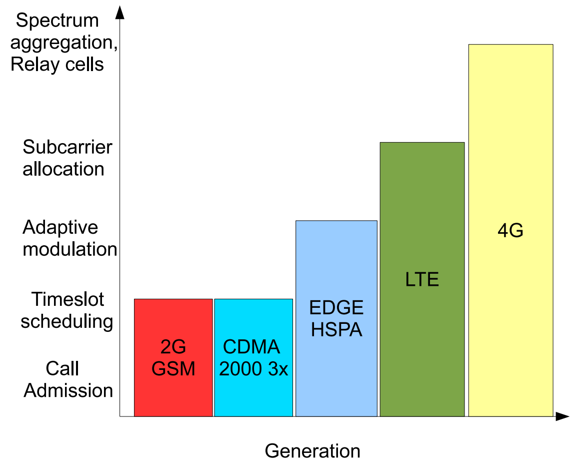


Figure 2.2: Evolutionary Path of Radio Resource Allocation

Downlink performance targets	LTE Release 8	LTE-A (4G)
Peak data rate	100 Mbps	1 Gbps
Peak spectrum efficiency	5 bits/sec/Hz	30 bits/sec/Hz
Average user throughput	3-4 times Rel.6 UTRA	-
Cell edge user throughput	2-3 times Rel.6 UTRA	0.07 bits/sec/Hz/cell/user
Average cell spectrum efficiency	3-4 times Rel.6 UTRA	2.4 bits/sec/Hz/cell
Latency	10 ms	10 ms

Table 2.1: Table of reference for performance targets [6],[7],[12] according to 3GPP for urban area coverage

It is worth noting that LTE-A is a candidate proposal for IMT-A. Upon comparing Tables 2.1 and 2.2 it can further be observed that LTE-A has set higher targets for urban area coverage as compared to IMT-A.

Environment	Cell spectral efficiency	Cell edge user spectral efficiency
Indoor (indoor cells)	3 bits/sec/Hz/cell	0.1 bits/sec/Hz/cell/user
Microcells	2.6 bits/sec/Hz/cell	0.075 bits/sec/Hz/cell/user
Urban area coverage	2.2 bits/sec/Hz/cell	0.06 bits/sec/Hz/cell/user
High speed	1.1 bits/sec/Hz/cell	0.04 bits/sec/Hz/cell/user

Table 2.2: List of 4G performance targets for different environments according to ITU [13]

2.2 Closed-loop MIMO

Multiple input multiple output (MIMO) is a multiple antenna concept that can achieve spatial multiplexing in a wireless channel.

The essential idea of spatial multiplexing is to communicate more streams of bits per unit bandwidth without using additional transmit power. A theoretical proof of this capacity was first published in Telatar et. al [26]. A transceiver architecture called V-BLAST (Vertical-Bell Laboratories) which can achieve the multiplexing gains was presented by Foschini et. al in [27], [28].

Another option is to use the so called diversity gain with multiple antennas, wherein a single data stream is transmitted redundantly using the multiple antennas. A rigorous characterisation of the trade-off between diversity and multiplexing tradeoff was derived in [29] .

Various schemes to exploit the diversity gains, multiplexing gains and their trade-off have been proposed. An exemplary scheme is the Alamouti space-time coding for two transmit antennas and one receive antenna [30]. For other antenna configurations, schemes that achieve high diversity with high data rates have been shown in [31], [32]. The abovementioned approaches can be realised without any channel knowledge at the transmitter, which is referred to as the open-loop approach.

In contrast, a closed-loop approach using feedback is shown in Figure 2.3. The difference between the approaches lies in the availability of channel knowledge at the transmitter, made through a dedicated feedback channel. The benefit of channel knowledge at the transmitter was characterised in [33].

It is possible to operate in closed-loop in LTE. A closed-loop approach is shown in Figure 2.3 which conveys limited channel knowledge in feedback. This is done using three parameters : channel quality indicator (CQI), precoder matrix index (PMI) and switching the number of streams (rank) [34]. CQI and rank switching are useful for link adaptation and achieving a packet error rate target. PMI feedback is useful to enhance the multiplexing gains when the receiver employs a linear equaliser such as minimum mean squared error (MMSE) or maximum ratio combining (MRC). The PMI feedback is performed in conjunction with rank switching so that throughput gain is realised

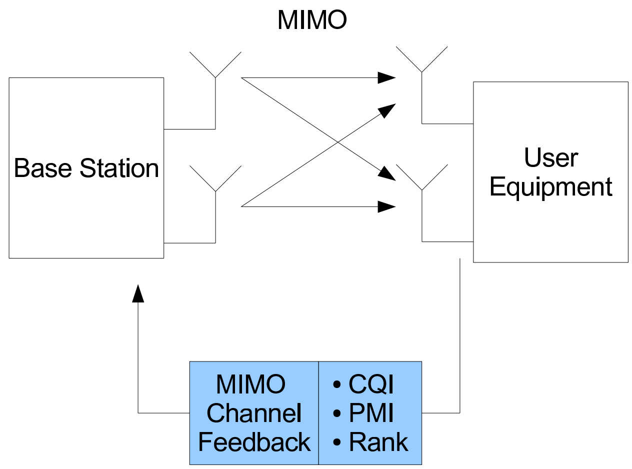


Figure 2.3: Closed loop MIMO

when there is rank switching.

The current proposal in 4G is to employ closed-loop MIMO in OFDM and thereby achieve a very high spectral efficiency for the downlink.

2.2.1 Enhanced MIMO concepts

Two enhancements to single user MIMO concepts have been discussed for 4G : multi-user MIMO and cooperative MIMO.

Multi-user MIMO is a single site MIMO concept, wherein a base station can transmit to many users in each frequency-time block by sptial multiplexing. The difference of multi-user MIMO to a conventional single user MIMO is that each of the many spatial data streams (in a time-frequency block) is intended for a separate user terminal. One simple illustration is a base station with 2 transmit antennas and 2 user equipments , each with 2 receiver antennas but with one spatial stream to receive.

Multi-user MIMO is already an optional transmission format in LTE, with latest LTE Release 10 [16] providing standards support to higher order MIMO upto 8 transmission antennas. More precoding options via extended codebooks (of bigger size) have also been included for multi-user MIMO in order to increase spectral efficiency.

Cooperative MIMO (also know as Coordinated multipoint, CoMP) refers to the concept of joint processing between adjacent base stations in order to achieve orthogonality in a time-frequency resource grid. In theory, cooperative MIMO can be realised through zero-forcing beamforming (for one receiver antenna) and block diagonalisation (a form of zero-forcing for two receiver antennas) [18]. A proof of concept real-time testing of cooperative MIMO has been conducted in the Easy-C project, which included various industry and academic partners. Some key enablers for CoMP are : distributed base station synchronisation, multi-cell pilots, high rate of channel feedback from users, and data plus channel information exchange between the cooperating base stations. The interface between the base stations is a X2 interface by using an optical fiber capable of transporting 1 Gbit/sec. The real-time testing of CoMP with two base stations each with 2 transmit antennas , and two user equiments each with two receive antennas has been performed in Easy-C trials and demonstrated in various conferences. In summary the results confirm the feasibility of cooperative MIMO approach in a OFDM cellular system using 20 MHz bandwidth and 500 m inter-cite distance [10] .

In this thesis, we do not specifically deal with enhanced MIMO concepts. The concepts and results in the thesis are shown only for single user MIMO. However our resource allocation and relaying results can be very well extended to the case of enhanced MIMO concepts. One straightforward change would be appropriately redefining precoder matrix indicators (PMI) in Figure 2.3 for the purpose of enhanced MIMO.

2.2.2 Test-bed description

The closed-loop MIMO approach as in Figure 2.3 has been implemented in OFDM for the real-time test-bed at Heinrich Hertz Institute in Berlin. Some of the issues that have been rigorously addressed in the test-bed are time and frequency synchronisation, phase noise correction and channel equalisation. In addition to real-time demonstrations, the test-bed also stores the MIMO channel matrices to quantify the results with offline simulations. In what follows, we outline the key hardware functionalities of the test-bed.

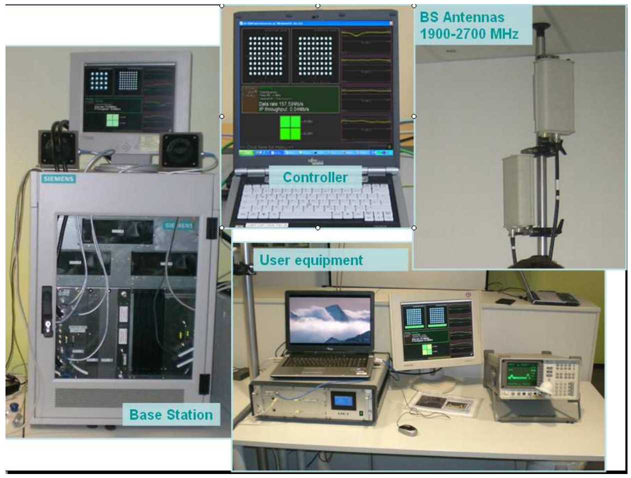


Figure 2.4: Main units of the experimental real-time test-bed

The experimental test-bed on MIMO-OFDM was planned according to the proposal of 3GPP-LTE air interface at the end of 2005. This was during the early stage of the LTE-standardization. The fully operational test-bed consists a user equipment prototype and a LTE-like base station. Figure 2.4 shows the base station and user equipment prototype. The test-bed uses cross-polarised antennas which are shown in the same figure to achieve high spectral efficiency in line of sight conditions. A detailed explanation of polarisation concept in the MIMO antennas has been presented in [119] and its potential benefits have been shown in [121]. The radio interface employs frequency division duplexing (FDD) such that both the downlink and the uplink transmissions are realised in real-time. A software 'controller' enables testing of the concepts such as MIMO rank switching and adaptive modulation. More details regarding the implementation of the test-bed is presented in [119].

As compared to the current definition of a resource block as 12 sub-carriers and 14 OFDM symbols in LTE standard, the test-bed implementation has been based on a definition of a resources block (RB) as 25 sub-carriers over 7 OFDM symbols. In principle this simply exchanges time granularity to frequency granularity. The following are the basic functionalities.

1. **Synchronisation** A coarse frequency estimation is performed at the user equipment before the fast fourier transform (FFT). For this purpose a special preamble is transmitted in the first time slot of a radioframe. The residual frequency offset between the base station and the user

equipment is estimated and corrected after the FFT by using the dedicated phase tracking pilots. Tests show the numerical values of the residual frequency offset to be in the order of a few 10 Hz in a static environment at a SNR above 10 dB. The OFDM frame synchronisation is based on another special preamble which uses a Schmidl-Cox correlator before the FFT at the user equipment. For the sake of robustness, the synchronisation preambles are transmitted with cyclic delay diversity (CDD) in order to sum up the power of the two transmit antennas without suffering from static beam-patterns caused by coherent transmission from the two antennas.

2. **Channel estimation**

High-quality channel estimation for 3G LTE is a challenge. Pilots are available only on a sparse grid in the time-frequency domain. On intermediate sub-carriers and time slots, channel interpolation has to be used. Since the same set of pilots are broadcast to all the users in the downlink, the theoretic solution would be to interpolate using all the pilots at each user equipment. However, simple calculations show this to be impractical with current digital signal processors (DSP). For example, in order to interpolate the LTE channel in 20 MHz bandwidth, one needs 1200 channel coefficients for two transmit and receive antennas from, say, each 4th sub-carrier being a pilot. For the optimal Wiener filter, this leads to computation of an interpolation matrix of size 1200×300. This process employs an 300×300 matrix inversion. Following this, the interpolation matrix is multiplied with a vector consisting of received pilots of size of 300×1. The complexity of matrix inversion is given as $300^3 = 27 \times 10^6$. Fortunately the inversion matrices can be pre-computed for certain SNR values. Therefore, the real constraint lies in the complexity of matrix and vector multiplication which is $1200 \times 300 = 3.6 \times 10^5$ and needs to be realized in real time. Furthermore, four times this complexity is needed in a 2×2 MIMO system (because of four antenna links). To do this, it requires roughly 4×10^9 real-valued instructions per second within a 0.5 ms time slot (assumed to be coherence time). This exceeds the capability of DSPs and therefore optimal interpolation cannot be performed. A simplification is localised interpolation.

Localised interpolation, as discussed recently in literature, relies on a subset of pilots around a desired data subcarrier. In [35],[36] it is shown that the loss in interpolation gain can only be a factor of 2 but reduces the computational burden by a factor of 20. A prerequisite for this gain is a

SINR estimation [35] so as to make the interpolation filter robust. A localised interpolation method has been implemented by applying local interpolation separately in groups of subcarriers, called a resource unit. This is done for each pair of transmit and receive antennas. Implementation is simplified by exploiting the fact that the interpolation matrices are independent of frequency indices when the local pilot grids are identical. Thus we adopt an approach of 'estimate and interpolate', wherein the SNR estimation is done first and the pre-calculated filter matrices are later applied based on the estimated SNR. In this way, the interpolation over the entire 20 MHz takes 95 μs for a 2×2 MIMO configuration, i.e. well within the coherence time of 0.5 ms.

3. **MMSE receiver**

The spatial separation of the data signals transmitted from two transmit antennas is performed in the field-programmable gate array (FPGA) on per sub-carrier basis. This step is after the FFT. Signal recovery is based on MIMO equalisation by using either the minimum mean squared error (MMSE) or maximum ratio combining (MRC) filter weights per sub-carrier. The weights are calculated in the DSP every 500 μs. Furthermore, the SNR for a group of data subcarriers is estimated and made available to the Viterbi-decoder. Due to the stringent timing requirements of calculating 1200 2×2 MIMO matrices every 500 μs, the process is split between 4 floating point DSPs (TI 6713) so that each processes 300 subcarriers.

4. **Latency Issues**

To separate the two spatially multiplexed data streams in the downlink a linear MMSE filter is used. Complexity of weight calculation scales linearly with the number of subcarriers but cubically with the number of antennas at one side of the link. The test-bed implements equalisation for 1200 used subcarriers and two spatial streams per subcarrier. Due to hardware constraints, signal processing has been divided among four DSPs and each is responsible for 300 subcarriers. Complex valued calculation is used for MMSE weight calculation. For the matrix inversion step in the MMSE, a closed-form expression of the inverse of 2×2 channel matrix is used mainly because it is less complex for 2 transmit and 2 receive antennas. A noise variance estimate that is needed for the MMSE calculation is obtained from the localised channel interpolation. Put together, coarse channel estimatation, channel interpolation and MMSE weight matrix calculation have been accomplished in about 0.3 ms in downlink and 0.45 ms in uplink, where the additional phase estimation is needed. The received signals are buffered after the FFT operation for 12 OFDM symbol durations. This buffering is prior to the MIMO-MMSE equalisation. The reason for this is to synchronise the input of MMSE weight matrices to the received signals.

5. **Antenna array beamforming** Commercial slant polarised antennas (+-45) developed by Kathrein [19],[20] with a fixed beam shape were used at the macro-base station. The proprietary antennas which are used by Fraunhofer HHI can operate within a frequency range of 1700 MHz to 2600 MHz, employing a 11 dBi antenna gain and a horizontal half power beam width of 60 which were suitable for measurements within a cell-sector. The measurements were conducted in the sector which contained the main beam from the antennas.

Each of the two antennas consists of four closely spaced and co-phased elements to produce transmit antenna array beamforming. No channel feedback is used for this purpose. In addition, user specific transmit beamforming by precoder feedback (PMI, channel feedback) from user equipment is supported at the baseband as mentioned previously in Section 2.2. The user receiver equipments use cross polarised (+-45) omni-directional antennas.

Chapter 3

Background Literature study

3.1 OFDMA

The problem of resource allocation in OFDMA is different to that of HSDPA. This is because OFDMA consists of sub-blocks of resource units called sub-carriers. As a result, a higher number of degrees of freedom can be realised through resource allocation in OFDMA. The channel knowledge of each of the subcarriers is available at the user equipment and can be informed to the base station. Thus downlink resource allocation can be done in a channel-aware manner. Importantly, on each of the subcarriers, adaptive modulation can be employed. The potential benefits of adaptive modulation in OFDM in single antenna systems was pointed out by Lawrey in [37].

In [37], the author also speaks about a technique of resource allocation called 'adaptive frequency hopping', in the sense that the users are allocated a set of subcarriers based on the channel state.

This technique is now referred to various names such as opportunistic resource allocation, opportunistic OFDMA or dynamic resource allocation. The idea is to dynamically vary the user-frequency map based on the channel states. Various proposals to utilise this technique continue to be a topic of current research.

In [38], dynamic resource allocation was used for an objective of maximising the minimum of the data rates across all the users, in what is popularly referred to as the max-min problem.

Other problem scenarios followed, notably the following two objectives : sum throughput maximisation or transmit power minimisation.

The transmit power minimisation problem was analytically solved in [39]. The sum rate maximisation problem is solved in [40] using two separate phases of subcarrier assignment and power allocation on the subcarriers. A more rigorous solution to the problem by treating both power allocation and subcarrier allocation jointly was derived by Agrawal et al in [41]. Both [40] and [41] use the relaxation approach, where they first relax the problem using Lagrangian relaxation. Interestingly, the conclusion that they reach is similar that allocating a subcarrier greedily to one particular user is optimal. Therefore the sum throughput maximisation is known as the greedy approach.

Strategies which introduce a desired amount of fairness to resource allocation were further proposed. A computationally efficient approach is the score-based scheduling proposal, which tries to allocate each user an equal share of the best resource blocks [42]. For characterisation of resource fairness, a well known metric is the proportional fairness which has been analysed in [43]. A proportional fair algorithm was first considered for wireless systems in [44] and proposed for high speed downlink packet access (HSDPA) systems. A main observation is that in a wireless environment where users are placed randomly, proportional fair algorithm results in a even distribution of resources [45]. OFDM versions of this algorithm have been presented in example works [46], [47], [48].

For applications with packet delivery deadline, it is natural to provide a minimum rate requirement. Some practical motivation and constraints in this context are discussed in [49].

A discussion of quality of service using exponential, logarithm and power utilities of data rate is presented in [50].

Another way of approaching quality of service issues in multi-carrier resource allocation is by weighting the subcarrier channel quality with the packet queue length [48]. This approach is known as delay-weighted scheduling and some further results with OFDMA can be found in [51]. However, the scheme does not propose a specific mechanism to handle situations of overload.

An interesting approach to providing user data rates is that of Pal et.al in [52]. They propose to maximise the number of satisfied users and provide solutions for high data rate (HDR) systems. The satisfied users are the users who are allocated a data rate equal to or greater than their demand. Their solution allocates time slots to users such that the demand for a maximum number of users has been met. We identify this 'maximum number of users' objective as the requirement for OFDMA resource allocation in this thesis.

For OFDMA, under the assumption of feasibility of all the user rate demands, various solutions have been proposed for allocation. A linear program is framed in [53] and [54] when the data rate requests are easily feasible. In the viewpoint that linear programs are difficult to solve, some low complexity heuristics for real-time implementation are also available. The heuristic in [53] uses a greedy allocation first and then a top-down descent which is also done greedily. In a similar approach [54] proposes a bottom-up ascent termed 'max-max' without an initial step of greedy allocation.

The so-called assignment problem, which is solvable in polynomial time using the Hungarian algorithm [55] is another possibility for resource allocation. The Hungarian algorithm has a particular advantage that it solves the $N \times N$ assignment problem optimally without a brute force search. In particular, the complexity of the Hungarian algorithm has been shown to be upper bounded by $O(N^3)$. This method has been used in [54] and [56]. Both works [54] and [56] approach this issue under the condition of feasibility. There are however subtle differences in how the algorithm is applied. [56] assume a pre-estimate of the number of subcarriers per user in a step they call resource estimate. This is done initially so that the Hungarian algorithm can be used. [54] however used a another assumption that the number of users is larger than the total number of subcarriers. This assumption removes the need for resource estimate and thus allows to directly apply the Hungarian algorithm.

The minimum data rate problem has been extended to a multi-cell multi-user scenario set up in [63]. Their basic idea is to enforce a minimum requirement in terms of number of resource blocks, which makes their approach similar to that of [56]. In addition to this, [63] points out that the minimum data rate problem is suitable for guaranteed bit rate services in the upcoming LTE networks.

In [59] and [60], the scenario is generalised to the case that some rate demands may be infeasible. In this context, a background literature on detecting the infeasible constraint set in a linear program is available in [58].

Generally speaking, infeasibilities occur when channel conditions to some users is adverse and yet many rate guarantees have to be met. For simplification, [59] and [60] modify the problem to maximising a common multiple of rate demands across all the users. A criterion is proposed in [60] for admission control based on the statistical estimate of the common multiple. However, a suitable algorithm for admission control and resource allocation based on instantaneous channel knowledge is not provided. In a general setting, an admission control mechanism may be needed to handle these situations more robustly. In this context, we refer to the paper in [62], where a simple admission control method is discussed for OFDMA. The method assumes a modulation and coding to be applied for the entire band and thus estimates the number of resource blocks. Moreover a resource allocation scheme is not provided when frequency dependent link adaptation is deployed.

3.2 Wireless Relays

3.2.1 Basic motivation for multihop

A wireless multiple hop model with n hops is shown in Figure 3.1. In this model, every radio receiver decodes the signal it receives from the previous hop and forwards it to the next hop. Let h_{ij} be the zero-mean quasi-static fading coefficient between any two hops i and j as a result of multipath fading characterised by a variance σ_{ij}^2 and n_i is AWGN noise at i^{th} radio receiver with variance N_0. The fading value h_{ij} does not vary for an arbitrary n time slots. The variance of the fading coefficient is in turn a result of the long-term fading effects, mainly path loss and shadowing. Let d_i be the Euclidean distance between two adjacent hops and there are n such wireless relay links placed in between a source and destination and thus $d_{sd} = \sum_i d_i$. The shorter wireless link between two relay nodes realise less path loss.

The path loss (PL) as a function of Euclidean distance d_{ij} w.r.t a reference point can be given by

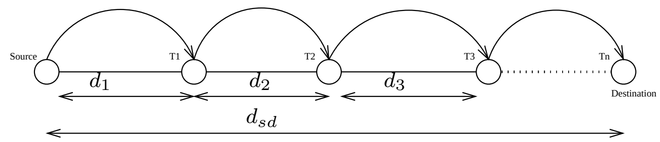


Figure 3.1: Multihop model with n hops

a simplistic model [21]

$$PL(d_{ij})[dB] = P_0 + 10 \times r \log_{10} d_{ij} \tag{3.1}$$

where r is the path loss exponent of the wireless channel model and P_0 captures both the propogation losses upto a distance of 1 meter from the transmitter and any shadowing loss on the propogation path to the receiver Note that $\sigma_{ij}^2 = 10^{-PL(d_{ij})/10}$. Further one may assume that P_0 is identical for all the links (i.e,. same transmitter and shadowing losses are applied on each radio link).

If P_{ti} is the average transmitted power from i^{th} node to its neighbor, then the average received power at the $i+1^{th}$ node is $P_{ri} = P_{ti}\sigma_{ii+1}^2$. Thus, if a minimum average received power of $P_{ri} = P'$ is required for coverage, then the transmit power P_{ti} would need to be adjusted to $P_{ti} = P'\sigma_{ii+1}^{-}2$. Note that for the direct link, uniformly dividing the source transmit power as $\frac{P_{t0}}{n}$ over n time slots would not improve it's overall received power after combining in a quasi-static channel.

In this case, the transmit power savings available using multiple hop communication can be calculated as

$$g = \frac{P}{\sum_{i=1}^{n} P_{ti}} = \frac{10^{P_0/10} d_{sd}^r}{10^{P_0/10} \sum_{i=1}^{n} d_i^r} = \frac{[\sum_{i=1}^{n} d_i]^r}{\sum_{i=1}^{n} d_i^r}. \tag{3.2}$$

It is easy to verify that g is always greater than one for $r > 1$, which means that relaying provides power-efficiency. We have presented some more results which show relaying to be energy-efficient in [115]. These gains can be understood to be one of the basic motivations for relaying. The assumption of applying same shadowing loss (P_0) on relay links as compared to the direct link is infact pessimistic because in principle the benefit from relays is to conveniently place them and get around severe shadowing as in [25]. Thus the power gain from relaying can be even higher than g by smart relay deployment concepts.

Few more motivations for relaying are discussed below :

- They can be a cost effective means to improve the reliability of communication. Studies in [64] suggest that the cost of a relay can be up to $\frac{1}{3}^{rd}$ of a base station and still provide profits to an operator. A similar conclusion has been made based on cost analysis and system level studies in [11].
- Relaying is a solution for link connectivity issues in a cellular network. One example of network coverage problem in cellular networks is the outage problem, where service is not available to users.
- Recent MIMO technology can be well utilised using relays, as for example reported in [65].

Theoretic view point of relays :

The first theoretic study of the relay channel was conducted in [66] by van der Meulen. In his pioneering work, relaying was addressed as a form of multiple-access channels. More information theoretic bounds were developed by Cover and Thomas in [67]. They characterised single relay scenarios and provide upper bounds for capacity considering different strategies : facilitation, full cooperation and partial cooperation.

Over the last decade, there has been a renewed theoretic effort for investigating efficient relaying techniques as in [68]-[69]. At the same time, practical deployment scenarios and throughput efficient usage of relays have been investigated.

Two classifications of relaying are now widely made : decode and forward (DF), and amplify and forward (AF). Simple AF and DF schemes based on full repetition were analysed by Laneman et.al [70]-[72]. They derive rigorous bounds for outage probability based on AF and DF schemes. They present simple relay protocols considering half-duplexing constraint at the relays, wherein a relay cannot receive and transmit at the same time. This half-duplexing constraint is a practical harware problem and thus relays incur a capacity penalty (also known as the half-duplex loss). Their work also gives an insight that the channel between a source and a relay is important to obtain the benefits from simple repetition based relaying schemes.

For wireless channels, relaying can provide diversity gains in addition to the path loss gains, the effect of which is referred to as 'relay diversity'. This effect can be realised from placing relays well separated, which will provide an independent multi-fading loss (channel realisation) from each of the relays to the receiver. The channel realisations between a relay and the user receiver can be time varying. Therefore, even in the absence of channel knowledge at relay transmitter, outage probability can be reduced at the receiver end from combining the information which arrive from statistically independent channels. The schemes in [70]-[80] exploit these gains and further analyse these gains mathematically using statistical channel gain distributions such as Rayleigh fading channels.

Sendonaris et al [81], [82], provide a scheme to use relaying diversity gain to improve cellular uplink via user cooperation. They propose a cooperative protocol with code division multiple access, which could be applied to the CDMA cellular networks.

The case of wideband relaying was further considered by Maric and Yates [83]. They compare AF and DF for communication between a *single source* and *single destination* with large operational bandwidth and a large number of relays. It was shown that performance of AF depends on bandwidth and power optimisation and is not efficient for wideband transmission i.e, an optimal bandwidth has to be chosen. Interestingly, a similar conclusion was earlier made by Schein in [84] for a parallel relay network. In the case of DF relaying, [83] showed that it is efficient to use wideband transmission. Moreover, they also argue that employing a single relay out of a number of relays is optimal for wideband DF relaying.

Cellular relaying :

The case for relaying in cellular networks is mainly to extend coverage. One scenario which is frequently experienced is shadow fading, wherein obstacles between in wireless medium result in a sudden drop in signal power. In such cases, relaying can be used to pass 'around obstacles' as was argued in [25]. To present the case, the authors used amplify and forward relays in a narrow band channel and derive the outage probabilities which show relaying to be beneficial. Thus, these simple repeaters can improve signal power at the receiver. The repeater approach is currently the popular way of utilising relays in cellular systems.

Esseling [86] and Pabst. et al [85] discuss deployment concepts using fixed relays in cellular networks. The fixed relays, which are also known as infrastructure relays are assumed to be DF relays which are specifically deployed by an operator to address coverage issues.

The work by Esseling et al. in [86] shows analysis of fixed relay nodes in a Manhattan street-grid scenario. In their work, they propose possible MAC-extensions of IEEE 802.11, HiperLAN and WiMAX, and evaluate the improved throughput performance of the individual systems. A key transceiver feature in their investigation is usage of duplex time sharing, where the base station-to-relay and relay-to-user links operate on the same frequency but use different time slots. Relays have been shown to substantially extend the radio coverage of a base station especially in highly obstructed service areas. They also present a table of spectral efficiency results for a different number of relays based on round robin scheduling at relays.

More improvement in cellular performance was observed by Hu et al [87] by deploying relay nodes in a hexagonal cell layout and deploying frequency reuse (with a reuse factor of 4) between relays in different macro-cells. The relay channels are fixed using the spatial reuse distance and cell planning criteria. The basic idea is to counter the capacity penalty incurred because of duplex time sharing with frequency reuse in different macro-cells. They perform simulations with adaptive modulation and coding on a 5 MHz bandwidth, with key assumptions of flat fading channel and no power control.

Comparison between DF and AF relaying schemes with LTE parameters is performed in [91]. They utilise duplex time sharing between base station and relay and optimise the time sharing values for sum throughput maximisation. Results show that spectral efficiency for DF outperforms AF when many relays are active at the same time. The authors however do not specifically address the case for supporting multiple users at a relay.

More in depth system level studies based on simulations are conducted for DF relaying using channel models in the project report of Wireless World Initiative New Radio (WINNER) [11]. They present coverage results by considering multiple users at a relay, but use a fixed round robin resource allocation. Even in this sub-optimal scenario, relaying has been found to be beneficial. Their studies also show that outdoor to indoor coverage without a relay is a challenge in OFDM systems. The worst 5% of indoor users receive a signal to noise distribution of only -5 dB, well below the +3 dB received by outdoor users. In terms of cell throughput, relaying improves average cell edge throughput by 50%. The average cell throughput is increased by 10%. Therefore it can be concluded that relaying is very useful to improve indoor coverage in particular at the cell edges.

In this viewpoint, the papers [94] and [34] discuss DF relaying for future LTE-Advanced networks. The importance of DF relaying for peak data rates in future LTE-Advanced systems is shown by simulations in [94] and some relay architectures are proposed in [34]. An interesting result in [34] is that full frequency reuse between adjacent macro base stations substantially improves coverage in cells with DF-relays.

More recently Oyman [88] analyses the effect of multi-user scheduling at relays on the sum rate maximisation. All of the abovementioned studies have been performed based on channel models.

Measurement studies with indoor dedicated antennas and repeaters were previously conducted for HSDPA systems in indoor locations [90]. Using an HSDPA system with up to 5 channelisation codes, 4 MHz bandwidth per code, the measurements showed that repeaters provided an average throughput of above 2.2 Mbps for open office corridor environment when operating at an optimum repeater gain of 65 dB.

We note that for OFDM systems, the IMT-Advanced target is to set 0.1 bits/sec/Hz for indoor environments at the 5 percentile point of cumulative distribution function of the user throughput [13]. This corresponds to 2 Mbps in 20 MHz bandwidth. The 2 Mbps data rate targets have also been adopted in recent WINNER studies [11]. For indoor OFDM, link measurements using channel sounding were conducted in [89] for the cases of outdoor to indoor and indoor to indoor. These measurements were performed in 5.3 GHz carrier frequency and 10 MHz bandwidth. They show capacity results based on Shannon capacity formula. The observation is that relays are particularly useful to improve the minimum data rates, which corresponds to solving coverage issues in a cell. Another main conclusion is that power allocation between base station and relay does not provide significant gains, with only a maximum improvement of 0.4 bits/sec/Hz.

Relaying and resource allocation :

Mukherjee and Viswanathan demonstrate the impact of relaying with user scheduling in relays for single carrier code division multiple access (CDMA) cellular system in [92]. They let multiple relays simultaneously transmit in a single time slot. Based on this approach, the impact of the number of relays is studied for 1.9 GHz carrier frequency, 1 MHz bandwidth using COST231-Hata channel model and 37 dBm relay transmit power. They observe that there is a 70% gain in sum throughput for 4 relays.

The problem of sum rate maximisation by using multiple subcarriers between a *single* source and *single* destination is solved by [95]. They provide solution based on power allocation and subcarrier allocation.

A heuristic approach for sum rate maximisation in a multiple relay multiple user scenario in OFDMA is given by [96]. They however allocate one subcarrier to only one relay link and perform fixed time division of BS and relay time slots. The sum rate maximisation problem is solved by power allocation to relays and subcarries via a Lagrangian approach in [97].

The objective of minimising the sum transmit power along with satisfying rate requirements is considered by [98]. Proportional fairness in a two hop scenario for multiple users is heuristically solved by [99].

Resource allocation for sum rate maximisation with multiple sources, multiple relays but single destination is shown in [100].

Optimum time slot optimisation for multiple relays aiding a *single* source and destination along with Qos requirements is considered in [101].

Relaying with fairness requirements is addressed in [102] and [103]. While both works exploit activating many relays simultaneously, they use different approaches to fairness. [102] resorts to assigning a minimum number of subcarriers as a measure of fairness. The authors formulate an optimisation problem which maximizes the system capacity and relies on channel state information of the complete network at the base station. [103] modifies the proportional fair metric by weighting with user rate demands.

3.3 Problem statements

The first problem we consider in the thesis is OFDMA macro-downlink resource allocation to provide guaranteed bit rates.

To the best of our knowledge, the problem of OFDMA resource allocation for the use case of guaranteed bit rates has not been well addressed in the abovementioned literature study. This problem is relevant to a service mode called 'Guaranteed bit rates (GBR)' in LTE [14].

The main limitation of existing OFDMA approaches is that non-GBR solutions (e.g, sum rate, proportional fair), though well researched, can not be directly applied for maximising the number of GBR users. Existing GBR solutions such as in [63] provide guarantees in terms of the number of resource blocks (by neglecting the channel quality of the resource blocks), but disregard the user bit rate requirement.

The user bit rate requirements are infact key to this problem. Each user in GBR may have a fixed bit rate requirement, for example, 64 kbps for conversational voice using ITU G.711 codec.

Thus the primary objective we consider for the base station scheduler for such use cases is to *maximise the number of users* who are given their guaranteed bit rates. We therefore ask the question, how to maximise the number of users such that each user can be given a required bit rate of d Mbps ?

A specific functionality that we look to exploit for GBR is admission control. Thus we term our approach as 'joint admission control and resource allocation'. There are admission control approaches for LTE, such as [62] which provides a simple admission control metric but does not deal with resource allocation procedures.

Another issue is complexity. The problem can be well related to linear programming (as we show later). However we note that linear programming-based approaches to find infeasible constraints as in [58] can not be used in practice. This is mainly because linear programming solvers need to handle a large number of variables. This will further involve an enormous number of computations and has to be accomplished within a few milliseconds. Moreover, the resulting resource allocations will be fragmented (many users sharing a resource block) because the solution set of linear programs are not integers and is moreover not supported in LTE.

Thus we emphasise the need for simple heuristics which can be implemented easily and solved in real-time. It is also desirable that the resulting solutions are *integer* and show good performance. However, linear programming-based solutions can still act as an useful upper bound for performance.

The second problem we look in the thesis concerns DF relaying and resource allocation for DF relays. Studies in [11], [94] have shown the benefit of relaying in OFDM with channel models. However the choice of model and its parameters can influence the result. We thus wish to illustrate our relaying protocols using measured channels in order to gain insight on the real world benefits of relaying. Prior relay measurements have been done in works such as [89] where channel sounding measurements with an indoor relay are shown. Those measurements used only a single antenna at all transmitters and receivers. In contrast, we wish to show relay measurement results with multiple antennas at base station, relay and user equipment and with adaptive modulation in this thesis.

The third problem we wish to consider is multi-user relaying for guaranteed bit rates which is an open issue.

Relaying schemes such as in [102] and [103] target user fairness in OFDMA with multiple relays. A relaying scheme with multiple relays is also presented in [91]. A shortcoming in the above mentioned relaying schemes is that the resource allocation at the relays and active time of the relays are treated as separate processes.

We propose to address this limitation by realising an inter-working of the two processes. A relaying scheme has to take into account the channels and data rates of users in each relay cell and with multiple such relay cells being active at the same time. For maximisation of the number of users in this challenging scenario, an optimisation protocol for relaying is developed in this thesis. An admission control protocol can be implemented at the relays and the base station. We look at this possibility in this thesis.

In regards to indoor relaying, we note that for OFDM systems, the IMT-Advanced coverage target is to set 0.1 bits/sec/Hz for indoor environments at the 5 percentile point of cumulative distribution function of the user throughput [13] for 10 users in a cell. This corresponds to 2 Mbps in 20 MHz bandwidth which have also been adopted in recent WINNER studies [11]. We thus present relaying results for this data rate.

Chapter 4

Downlink resource allocation

4.1 Overview of quality of service

Centralised resource allocation can be performed in cellular systems to exploit the channel independence among many users, which is termed multiuser diversity. Exploiting multiuser diversity has been a popular aspect in recent systems such as 3GPP high speed packet access (HSPA). In 3GPP LTE, multiuser diversity can be exploited in each subcarrier using orthogonal frequency division multiplexing access (OFDMA). Blocks of subcarriers can be selectively allocated to a user based on the channel response. This is called scheduling. The LTE downlink can increase its sum throughput using this opportunistic resource allocation.

There is always a requirement for fairness because users are distributed at different distances and a user at the cell edge may never be served. Thus a base station can not serve a cell with a sole objective to maximise the sum throughput in the cell. To provide fairness, a base station can effectively use two kinds of informations about a user : channel quality information and quality of service requirements. Thus, quality of service and fairness are interlinked issues, where fairness is used as a tool for providing quality of service.

To assist in scheduling, quality of service information can be utilised from application specific attributes such as service priority, deadline and packet error rate which are set specific by the network. Application here refers to the class of traffic such as voice, streaming video or file download which are broadly classified as real-time or non-real-time. These applications can be supported in one of two ways : guaranteed bit rate (GBR) or non guaranteed bit rate mode [14]. A service offering can thus be made for a user in one of the two modes.

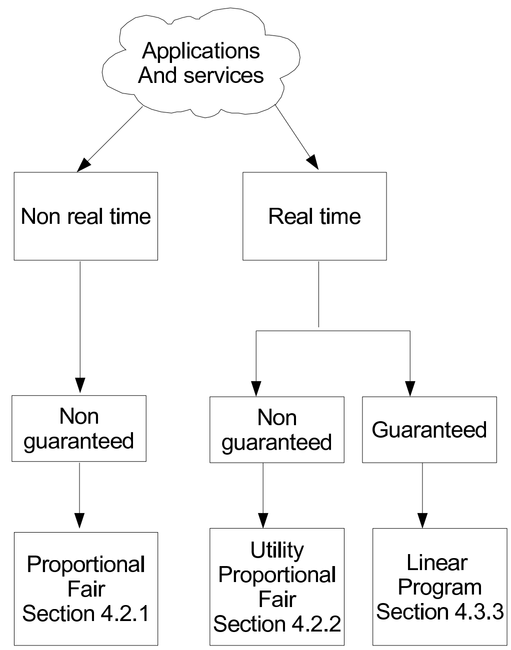


Figure 4.1: Mapping between service category and scheduling fairness

In Figure 4.1 we show a mapping between the type of service offering and the fairness policy.

For each classification shown in the figure, our approach is to use OFDMA scheduling and efficiently use the frequency selectivity of the channel. As a motivational example, the frequency response of a frequency selective channel is shown for two different spatial positions in Figure 4.2. This example is shown from the outdoor measurements in the site reported in [116]. The measurements site is of dimensions approximately 300 $m \times 160$ m and the measurements were collected from different spots in the cell using the test-bed described in Section 2.2.1. The frequency response is shown for two locations in the cell and from the same base station. Two properties are evident : a) the channel quality varies over the subcarriers and b) the channel frequency response varies over the spatial positions. Thus users who are in different cell locations would experience different channel frequency responses. In what follows, we show the benefit of resource allocation for the non-guaranteed and guaranteed bit rate modes. For comparison to some baseline resource allocation schemes, we refer

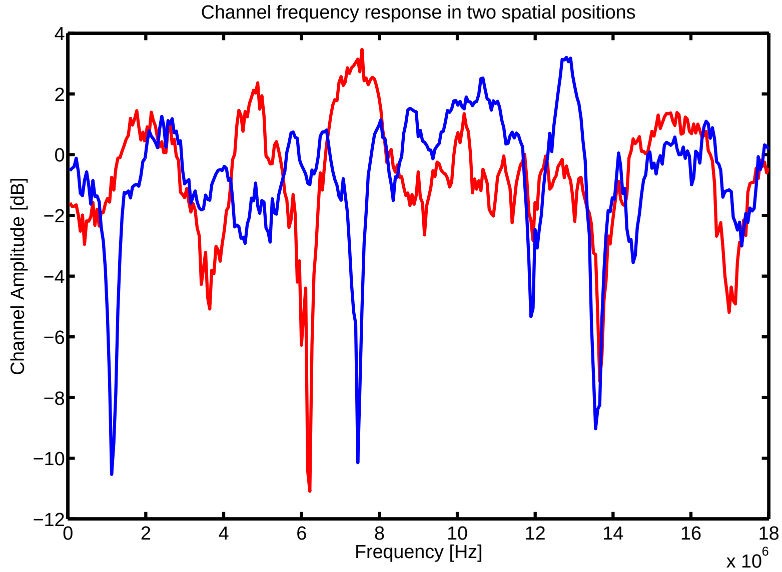


Figure 4.2: Normalised channel frequency response at two spatial positions. Obtained from a measurement outdoors in 18 MHz bandwidth and using 336 downlink reference symbols

our results to the sum throughput maximisation (greedy) and round robin (fairness) schemes. Furthermore, we show that schemes using the non-guaranteed bit rate mode are ill-suited to the guaranteed bit rate. With this in mind, we propose specifically designed resource allocation schemes that are suitable for the guaranteed bit rate mode. The main contributions of the chapter are a) the analysis of all the schemes classified in Figure 4.1 and b) the derivation of a number of novel OFDMA methods for the guaranteed bit rate mode. All the simulations use stored set of channels that are obtained from measurements using the test-bed described in Section 2.2.1. The channels are first normalised for unit power (by summation over all the locations) and then an average signal to noise ratio (SNR) of 15 dB is used.

The following premise used throughout this chapter.

- Resource block : A set of subcarriers in a group of OFDM time symbols in a transmit time interval is defined as a resource block. With a slight change in LTE parameters, we define a resource block as 25 adjacent subcarriers in 7 OFDM symbols which last for a total duration of 0.5 ms. A data symbol is mapped to the elementary grid of a data subcarrier in an OFDM symnol. Discarding the pilot tones in the resource block, there are 144 data symbols in a resource block of 0.5 ms duration. Specifically, for 20 MHz bandwidth there are 48 resource blocks. For scalability, we may however generalise the number of resource blocks to be M and each resource block to consist of L subcarriers for a duration of T OFDM symbols.

- Modulation and coding scheme : Downlink reference symbols are transmitted by the base station in the frequency domain for the purpose of channel estimation at the user equipment. There are 336 such reference tones in 20 MHz bandwidth. After channel estimation, a user equipment can feedback the quantised channel matrix of each reference tone to the base station to enable closed loop MIMO and adaptive modulation. However, channel magnitude and phase feedback for MIMO-OFDM can overly consume the uplink bandwidth. A limited feedback scheme is used as follows : the user equipment computes the signal to noise ratio (SNR) per spatial stream using the reference symbols after applying a minimum mean squared error filter. A one-to-one mapping is used between the SNR estimation and the modulation and coding scheme (MCS) based on a look up table. Thus only the MCS level per spatial stream corresponding to each reference tone is informed to the base station by the user equipment, instead of precise channel values. The MCS level can be applied on all data subcarriers around the reference tone. For even more reduction in the feedback data, the MCS value is computed on a resource block basis instead of a reference tone basis. To meet quality of service requirements, the MCS levels per resource block are obtained from a look up table that have been designed not to exceed a packet error rate. The WINNER proposal for modulation and coding set [109] is shown in Table 4.1 for a packet error rate of 10^{-2}. In practice, the packet error target can vary and be set specific to the application as recommended in [14]. It is further noted that for the presentation of results, channels measured in interference-free scenario were used. However an interference level can be introduced and average SINR (signal to interference noise ratio) value may be applied to these channels depending on the level of interference that is expected.

- Uplink feedback : The uplink feedback regarding channel quality is performed once in a radio frame , i.e, once per 10 ms radio frame. This feedback rate has been found to be sufficient for reliable downlink scheduling in real-time tests [116], [119], [120]. These tests were conducted in low mobility scenarios (up to 30 Kmph) and very low mobility scenarios (3 Kmph). Doppler

Modulation	Code rate
No transmission	No transmission
BPSK	$\frac{24}{36}, \frac{24}{40}, \frac{24}{44}, 0.5$
QPSK	$\frac{24}{28}, \frac{24}{30}, \frac{24}{32}, \frac{24}{34}, \frac{24}{36}, \frac{24}{40}, \frac{24}{44}, 0.5$
16-QAM	$\frac{24}{30}, \frac{24}{32}, \frac{24}{36}, \frac{24}{40}, \frac{24}{44}, 0.5$
64-QAM	$\frac{24}{26}, \frac{24}{28}, \frac{24}{30}, \frac{24}{32}, \frac{24}{36}, \frac{24}{40}, \frac{24}{44}, 0.5$

Table 4.1: Modulation and coding set for PER 10^{-2}

effects and schemes to improve robustness in high mobility are therefore not considered in reporting MCS levels.

- Uniform power allocation : The total transmit power at the base station is divided equally among all the subcarriers. Thus, if P is the total transmit power per OFDM time symbol, the transmit power per subcarrier per OFDM time symbol becomes $\frac{P}{LM}$. Even though an optimisation problem can be framed by treating the transmit power per subcarrier as a variable, we apply uniform transmit power allocation among subcarriers. In effect, this approach can utilise the multiuser diversity in the received power through proper scheduling at the transmitter. In comparison, literary work [41] reports minimal performance change between uniform and non-uniform power allocation for the objective of OFDMA sum rate optimisation. A particular drawback of the non-uniform power allocation, apart from the higher computational complexity, is that more feedback is required. Thus we optimise the fairness objectives in Figure 4.1 simply based on uniform power allocation but through proper scheduling of resource blocks at the base station.

4.2 Non-guaranteed bit rate allocation

In non guaranteed bit rate mode a network offers service to users without negotiating a specific bit rate level. The basic assumption is that the service flows in non guaranteed mode can accommodate congestion related packet drops. The selection of this mode can be initiated by a user, the network or even an application. In this mode there is no obligation for a scheduler to deliver a specified number of packets within a time duration. Thus a base station scheduler can maximise a utility function based on time slot and resource block sharing. There are a variety of utility functions depending on an operator defined notion of fairness as previously discussed in the literature study in Chapter 3. Here we consider a popular approach called the proportional fairness [43] for OFDMA. There are two versions of it : rate proportional fairness and utility proportional fairness. Optimal and suboptimal schemes for the two methods are also presented.

4.2.1 Rate proportional fair

The rate proportional fair method is proposed for non-real-time flows. Non-real-time flows are service flows without a strict time deadline. Because of the flexibility in time deadline, absolute bit rate targets are not needed in the short-term and hence these flows can be supported in non-guaranteed bit rate mode. An example is a file download from the Internet, for which the service quality is perceived by users in the order of few seconds, while the transmit time interval is in the order of few milliseconds. Moreover, there is no strict time deadline to deliver the packets within some seconds. Thus a scheduler can deliver the packets in a flexible time duration under a fairness scheduling rule. Proportional fairness is a rule which looks for a fine balance between the fairness and efficiency in scheduling. The ultimate target of the proportional fair method in cellular wireless is to serve each user on its own relative channel peak, so that both efficiency and fairness are achieved.

In OFDMA systems, proportional fairness can be achieved by utilising both time slots and frequency subcarriers; unlike CDMA systems where only time slots are utilised. A possibility in OFDMA from the perspective of downlink is that fairness can be achieved by utilising an independent MCS value in each frequency resource block; unlike the CDMA systems where a single MCS level is loaded in an entire time slot. Below we describe algorithms that can exploit this advantage.

4.2.1.1 Optimal algorithm

Assume a frequency selective channel, where a specific MCS is supported on each subcarrier. Suppose u_{kf} is the MCS that can be supported for user k in subcarrier f and x^t_{kf} denotes the resource allocation in terms of time slots that is allocated to user k on subcarrier f. The objective of the proportional fair scheduler is to maximise the log-sum-of-rates [1] subject to resource constraints that is denoted as

$$\max T \sum_{k=1}^{k=K} \sum_{f=1}^{f=LM} log[x^t_{kf} u_{kf}] \tag{4.1}$$

$$\sum_k x_{kf} = 1 \quad \forall f \tag{4.2}$$

$$0 \leq x_{kf} \leq 1 \tag{4.3}$$

From the definitions of MCS and resource blocks, we know that the same MCS is applied for the L subcarriers in a resource block. Thus we employ this information for computation.

By denoting a resource block using variable m, there are M such resource blocks of L subcarriers each. Thus u_{km} is the MCS that is loaded for user k on resource block m. Replacing $\sum x^t_{kf}$ with x_{km}, we get,

$$\max LT \sum_{k=1}^{k=K} \sum_{m=1}^{m=M} log x_{km} u_{km} \tag{4.4}$$

$$\sum_k x_{km} = 1 \quad \forall m = \{1..M\} \tag{4.5}$$

$$x_{km} \geq 0. \tag{4.6}$$

Thus in (4.4)-(4.6), we drop the time index t and denote x_{km} as the amount of subcarrier allocation to user k out of each resource block m. We have also dropped the contraint $x_{km} \leq 1$ because this condition becomes implicit w.r.t (4.5) and (4.6).

The objective in (4.4) is convex in terms of x_{km} and the constraints are linear and therefore this is a convex optimisation problem.

Below we show the optimality condition of the variables x_{km} using the Lagrangian multipliers

[1] is a utility function to model satisfaction. No physical units are attached to utility functions [24]

approach. For this purpose, let the multiplers λ_m be introduced for the equality constraint in (4.5) and the multiplers β_{km} for the inequality constraint in (4.6).

Upon taking the derivate of Lagrangian function of (4.4)-(4.6), equating it to zero and using the fact that $\beta_{km} = 0$, for $x_{km} > 0$, the following condition is written

$$x_{km} = \frac{1}{\lambda_k - \beta_{km}} - \sum_{j \neq m} x_{kj} \frac{u_{kj}}{u_{km}}, \tag{4.7}$$

$$\text{where} \quad \beta_{km} > 0, \text{if} \quad x_{km} = 0, \tag{4.8}$$

$$\beta_{km} = 0, \text{if} \quad x_{km} > 0. \tag{4.9}$$

The optimal values of x_{km} can not be obtained trivially from (4.7). The optimal values of λ_k first need to be obtained, which however is dependent on all other x_{im}, for $i \neq k$ because of the constraint (4.5). Thus finding the optimal solution via Lagrange multipliers requires a multi dimensional search to update all the multipliers based on the constraints in (4.5) and (4.6). Methods for efficient computation exist, for example the interior-point method in convex optimisation [105]. An efficient implementation of the interior-point techniques with some advancements is available in Matlab-based tool such as *fmincon*. We use this function to compute the optimal solutions. We remark that optimal solutions would result in sub-resource block scheduling to users (i.e resource block scheduling with fractional values) which is not part of LTE. However optimal solutions serve as a useful upper bound to scheduling performance.

4.2.1.2 Suboptimal algorithm

As can be deduced from the optimality conditions above, the optimal solution for proportional fairness is not easy to obtain in OFDMA. Thus, suboptimal resource allocation approaches for fast and efficient maximisation of the objective in (4.4) are needed. Here we present a new scheme which extends the W-CDMA version to include frequency subcarriers of OFDMA. To develop suboptimal algorithms, we treat the problem as a selection problem of user-resource block pairings and thus look for binary integer (0 or 1) solutions of the variables x_{km}. Specifically, we wish to use the MCS information of user-resource block pairing. A formal description of the suboptimal algorithm is presented below, which we term the gradient search scheme. We have presented part of these results in [120].

1. Initiate a resource block pool, V, consisting of all the resource blocks. Initiate $x_{km} = 0 \quad \forall k = 1..K, \forall m = 1...M$.

2. Calculate weight factors for users as

$$w_k = \sum_m x_{km} u_{km} + \delta, \tag{4.10}$$

   where δ is a regularisation factor to set a non zero weight value w_k.

3. Select the resource block and users pairing according to $(k^*, m^*) = \arg\max_k \max_{m \in V} w_k u_{km}$, i.e, search for the best combination of user index and resource block. Set $x_{k^* m^*} = 1$. Remove the resource block m^* from the pool V. Repeat until V is empty.

The gradient search scheme allocates a resource block m to user k if the pairing (m, k) corresponds to the maximum bit rate gradient. To perform the selection to a user, a resource block sort is needed for each user in each iteration based on the MCS values. This effectively means that the resource blocks are not allocated in some fixed order.

Simulation results for the following schemes : a) Round robin scheduling b) Suboptimal gradient search and c) Optimal proportional fairness are presented in Figure 4.3

4.2.2 Utility proportional fair

The utility proportional fair method [2] is used for real-time flows in the non guaranteed bit rate mode. Packets belonging to the real-time service flows need to be scheduled within a time deadline. An example is a multi media streaming video, for which a user's experience is affected by delay

[2] we call this utility proportional fair because it targets proportional fairness of a utility metric. Note that proportional fairness is itself a utility function

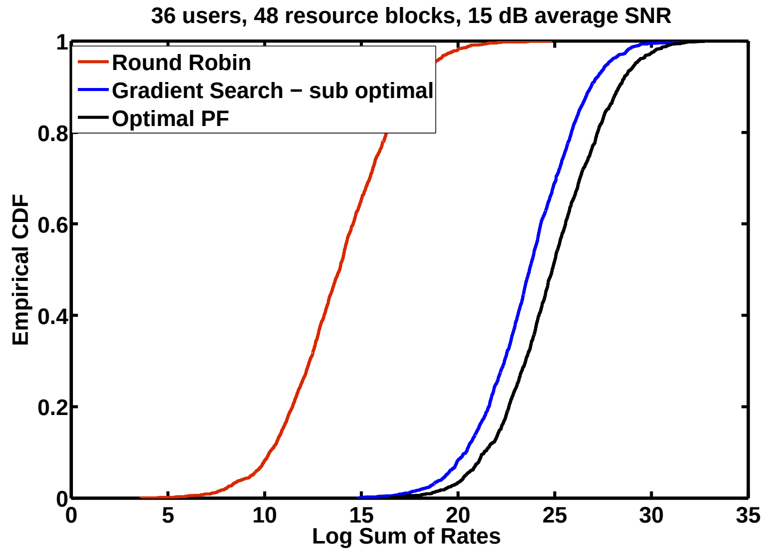


Figure 4.3: Performance comparison of resource allocation schemes in terms of the proportional fair objective.

in packet arrivals, which are called jitters. Thus for these kind of applications, a short term bit rate target, denoted as Ω_k is applicable in scheduling to user k. It may not always be possible to meet the requirement of each user when there are too many users. Thus in the case of congestion in the network, a scheduler has to evaluate the benefit of a rate allocation to a user. The benefit can be modeled using a utility function wherein, above a minimum data rate, denoted as Ω_{min}, the usefulness of the data rate increases steadily. Below this minimum data rate, the service quality becomes unacceptable. Importantly, there is not a significant change in the perceived service quality for data rates from zero until Ω_{min}. This minimum bit rate is thus used as a reference by the scheduler to measure the quality of service to a user. Similarly, above the target bit rate Ω_k, there is no significant change in the perceived service quality. To model this approach, we make use of sigmoidal logistic function. The sigmoidal function is described by three regions : a) minimum bit rate Ω_{min} , b) inflection point $\tilde{D}$, where the function changes convexity and c) the saturation bit rate, denoted as Ω_k, which is also the target bit rate for the scheduler.

One can approximate linearity in the region between Ω_{min} and Ω_k, in which case the relationship to the inflection point $\tilde{D}$ becomes

$$\tilde{D} \approx \frac{\Omega_k + \Omega_{min}}{2}. \tag{4.11}$$

The sigmoidal utility [3] is defined using the value of $\tilde{D}$ as

$$\Psi_k = [1 + \exp\left[s[\tilde{D} - (\sum_m x_{km} u_{km})]\right]]^{-1}, \tag{4.12}$$

where we recall that $\sum_m x_{km} u_{km}$ is the data rate that is served to a user k.

The parameter s in (4.12) characterises the slope of the function and can be set to achieve a certain value of Ψ_k at $\sum_m x_{km} u_{km} = \Omega_{min}$. Let $\Psi_k = 0.1$ for $\sum_m x_{km} u_{km} = \Omega_{min}$. Based on this value, the parameter s can be found to be $\frac{\ln 9}{\tilde{D} - \Omega_{min}}$.

Now we wish to perform OFDMA resource allocation using the sigmoidal utility function as a tool. The maximisation problem for K users is written as

$$\max \sum_{k=1}^{k=K} log \Psi_k \tag{4.13}$$

$$\sum_k x_{km} = 1 \quad \forall m \tag{4.14}$$

$$0 \leq x_{km} \leq 1. \tag{4.15}$$

It is noteworthy that we use an objective of log-sum-of-sigmoidals using the function (4.12) instead of sum of sigmoidals. The goal is to exploit the fact that the sigmoidal function is a log-convex function. Therefore, any maximal value of the objective in (4.13) is also the global optimal. To obtain the maximal values of x_{km}, we may again use an interior point method in convex optimisation.

For a suboptimal scheme, we use a gradient approach as before in Section 4.2.1.2 based on the gradients of utilities Ψ_k.

[3] No physical units are attached to utility functions [24]

1. Initiate resource block set V consisting of all the resource blocks. Initiate $x_{km} = 0 \forall k = 1..K, \forall m = 1...M$.

2. Calculate weight factors for users in resource block m as

$$w_k = \Psi_k + \delta, \tag{4.16}$$

where $\Psi_k \geq 0 \quad \forall k$ is obtained using (4.12) for given values of Ω_{min}, Ω_k and s. δ is a regularisation term to set a minimum value of w_k.

3. Select the resource block and user pairing according to

$$(k^*, m^*) = \arg\max_{k} \max_{m \in V} w_k(\Psi_k(x_{km} = 1) - \Psi_k(x_{km} = 0)), \tag{4.17}$$

i.e, search for the best gradient of log-sigmoidal using user index and resource block. Set $x_{k^*m^*} = 1$. Remove the resource block m^* from the pool V. Repeat until V is empty. Simulation results are presented in Figure 4.4

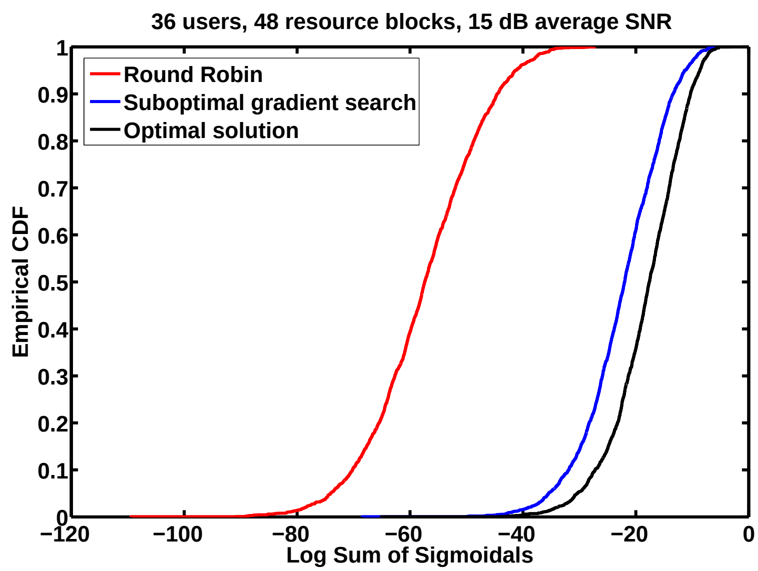


Figure 4.4: Performance comparison of resource allocation schemes in terms of log-sum-of-sigmoidal utility function. Target rate Ω_k=2.25 Mbps. Ω_{min}=$\frac{\Omega_k}{2}$ Mbps

These results of utility PF have shown the case when two data rates Ω_{min} and Ω_k characterise the quality of service. An intuitive understanding of utility PF is that, the benefit of allocating a rate less than Ω_{min} is low and Ω_k acts as an upper limit. The limiting case of $\Omega_{min} = \Omega_k$, will be addressed in the following sections as a separate problem, in what is termed as guaranteed bit rate.

4.2.3 Summary

The two schemes for non-guaranteed bit rate mode have been shown. In addition, two novel suboptimal techniques based on maximum gradient have been derived. The optimality conditions are presented. Numerical results show performance within 5% of optimum and 71% gain over round-robin. In all cases the greedy scheme is non-viable for QoS purposes. The next section will consider the guaranteed bit rate mode.

4.3 Guaranteed bit rate allocation

This section is the main contribution of this chapter. Here we propose resource allocation methods for the guaranteed bit rate (GBR) mode [14]. The guaranteed bit rate mode works through a bit rate negotiation between the network and a user. The network has an obligation to fulfill the negotiated bit rates. The provisioning of a specified bit rate is understood to maintain an end-to-end connection with the user. A network may have to cater too many GBR users, in which it case the radio interface may encounter a capacity bottleneck. A bottleneck scenario may arise because a base station is expected to transmit the guaranteed amount of data to all users but with a given time duration, a limited bandwidth and a total power constraint. The task of a base station scheduler is thus to appropriately schedule the resources to minimise the connection drops. One approach to solve the problem is to use a time slot based admission control. In the following sections, we describe this process of admission control and resource block allocation.

4.3.1 System Model

The following system model is assumed for GBR mode as shown in Figure 4.5. Feedback from a user equipment is shown in the figure, in which the user equipment can select a quality of service

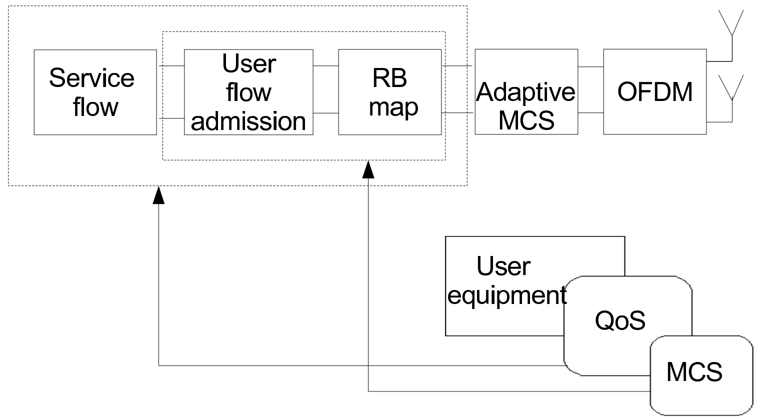


Figure 4.5: Functional description of downlink scheduling for guaranteed bit rate.

and inform the base station of the choice. In addition, the user equipment provides feedback about the estimated channel quality. As in Section 4.1, we apply the feedback scheme in which the user equipment informs MCS level in each resource block. For negotiating a connection with the network, a user equipment is assumed to be aware of the QoS options of the network. For example, a content may be available in different qualities at the server and the user equipments selects one among them. Assuming this mechanism, in Figure 4.5, we refer service flow as the bit rate level of the content which is requested by a user equipment. We use admission control on a transmit time interval (TTI) basis on the user's service flow before performing resource block allocation. One user equipment may require several service flows, in which case the base station scheduler may selectively admit certain flows of a user. Thus admissibility of a user's flow and resource block scheduling depends on the MCS feedback and the QoS request from a user.

4.3.2 The need for admission control

Admission control is a mechanism which can be implemented in a network to avoid any quality of service degradation.

In a OFDMA cellular network with base station scheduling, we propose two admission controls : an outer and inner admission control. The outer admission control is implemented at a sessional level at the gateway and implements the traffic contract in the network. The inner admission control is implemented at the base station level before scheduling radio resources to users at the transmit time slot. Thus the inner admission control infact represents a user pre-selection for the scheduling process at the base station.

We however note the possibility that the inner admission control can be used to steer the decision making of the outer admission control even before a traffic contract is negotiated with a user. The inner admission control can then effectively inform the outer admission control whether radio resources are sufficient at that base station to grant that user's guaranteed bit rate connection request. In practice however, the outer admission control would also be driven by a host of other factors such as charging policies (whether an user is premium user) and backbone network throughput.

In the spirit of the above argument, the inner admission control can thus be used to minimise the dropped users (i.e maximse the served users), in case the outer admission control has already (even perhaps erroneously) admitted too many users. Intuitively, we note that the task of the inner admission control is therefore to admit users for the scheduling process based on the relative channel conditions in order to handle infeasibilities.

Generally speaking, infeasible situations may arise in GBR mode because of adverse channel conditions and too many users. In the case of too many GBR rate requests, the likelihood of infeasibility increases because the downlink is bandwidth limited. Thus a mechanism may be needed to handle arbitrary user rate requests. In what follows, we refer to the inner admission control just as 'admission control' purely from the perspective of radio resources at a base station and assume that an outer admission control is already existent.

Consider a example scenario as follows. Two users A and B require a similar target bit rate for their applications. The two users request the rate to the network using the feedback channel. User A is in better channel conditions and places a straightforward rate request while B is in relatively worse channel conditions but greedily asks for a higher rate. Thus without a proper mechanism in place, guaranteed allocation for A is degraded by B's request, even though A can be served with the target rate. In short, a network must not let a few rate requests to dominate the rate allocations for the users of the network.

Thus we aim to maximise the number of satisfied users, i.e the number of users who are allocated the requested rate (the guaranteed rate). We achieve this objective by using admission control along with resource block allocation on a time slot basis.

4.3.3 Optimal solution

Assume a frequency selective channel with u_{km} as the MCS level supportable to user k on resource block m. Let Ω_k be the guaranteed number of bits that has to be provided to user k's flow within a time duration of T symbols, (termed as the guaranteed bit rate). The problem is represented in the linear programming form as

$$\max LT \sum_{m=1}^{M} \sum_{k=1}^{K} u_{km} x_{km} \tag{4.18}$$

$$\text{s. t ,} \quad LT \sum_{m=1}^{M} u_{km} x_{km} \geq \Omega_k \quad \forall \{k = 1, 2.., K\} \tag{4.19}$$

$$\sum_{k=1}^{K} x_{km} = 1 \quad \forall m = \{1, 2, .., M\} \tag{4.20}$$

$$0 \leq x_{km} \leq 1 \quad \forall k, \forall m, \tag{4.21}$$

where x_{km} represent the fraction of subcarriers that is allocated from resource block m to user k. The constraint in (4.19) specifies that the scheduled bit rate to user k must be greater than the target Ω_k. (4.20) is the resource constraint of L subcarriers per resource block. If the variables x_{km} should only be 0 or 1, the problem is an integer programming (IP) problem, for which the optimal solution is difficult to obtain. In the allocations, more than one resource block can be allocated to a user and in each of the resource blocks a different MCS level can be loaded. We refer to the problem in (4.18)-(4.21) as the multi-user loading problem.

The bit rate demands in (4.19) may be initiated by user equipments via feedback as shown in the system model in Figure 4.5. These bit rate requests are used as the guaranteed bit rate values by the network. All of these bit rates may not be admissible to the base station. Therefore, the constraint set in (4.19)-(4.21) can be infeasible in which case the challenge is to identify the best feasible subset.

The problem is to identify and remove the infeasible constraints so that resource allocation can be efficiently performed for the remaining constraints. This process is not trivial because in most cases there is no prior knowledge about the infeasible constraints. In this situation, we re-formulate the objective to satisfy as many constraints as possible while performing a secondary objective of sum rate maximisation. The objective functions for multi-user loading are re-defined as

$$\max \sum_k a_k \quad \text{s.t,} \quad a_k = 1, \quad \text{if} \quad LT \sum_m u_{km} x_{km} \geq \Omega_k \quad \forall k \in \{1, 2, .., K\}, \tag{4.22}$$

$$\max \sum_k r_k \quad \text{where,} \quad r_k = LT \sum_m u_{km} x_{km} \quad \forall k \in \{1, 2, .., K\}, \tag{4.23}$$

where Ω_k is minimum data rate requirement of user k. A brief numerical example is shown to describe the working of multi-objective optimisation in (4.22)-(4.23). Suppose there are 4 users and each user demands 1 Mbps. Consider two resource allocation policies : Policy A which allocates a rate tuple in Mbps as $(1.7, 1.5, 0.7, 0.6)$ and another policy B which allocates $(1.1, 1, 1, 0.3)$. Then on account of the user objective in (4.22), policy B is better than policy A. If there exists another policy C which allocates $(1.1, 1, 1, 0.4)$, policy C is more efficient than policy B because of the better network objective (4.23).

One way to achieve the maximisation in (4.22)-(4.23) is to search for binary integer solutions of $x_{km} \quad \forall k, \forall m$ with full enumeration of user and resource block pairings. In case the maximising solution for the primary objective $\sum_k a_k$ is not unique, then the solution which additionally maximises the objective $\sum_k r_k$ is optimal. Full enumeration consists of K^M entries and this complexity becomes prohibitive for a high value of M. Therefore, we look at low complexity approaches for the search.

4.3.3.1 Joint admission control and allocation

Focusing on the primary objective in (4.22), we begin by introducing binary variables a_k for admission control and re-write the problem as

$$\max \sum_{k=1}^{K} a_k \tag{4.24}$$

$$\text{s. t} \quad LT \sum_{m=1}^{M} u_{km} x_{km} \geq \Omega_k a_k \quad \forall k \in \{1, 2, .., K\} \tag{4.25}$$

$$a_k \in \{0, 1\} \tag{4.26}$$

$$\sum_{k=1}^{K} x_{km} \leq 1 \quad \forall m \in \{1, 2, .., M\} \tag{4.27}$$

$$0 \leq x_{km} \leq 1, \tag{4.28}$$

where the equality in (4.20) has been changed to inequality in (4.27). Thus solutions for a_k and x_{km} need to be found and the two problems in (4.24)-(4.28) are coupled. Hence we term this joint admission control and resource allocation. Now we relax the variables a_k to be real numbers between 0 and 1. It can be deduced from (4.25) that $a_k = \Omega_k^{-1}[\sum_{m=1}^{M} u_{km} x_{km} - s_k]$, where $s_k \geq 0$ are slack variables so as to replace the inequality in (4.25) with equality. By further setting $s_k = 0 \quad \forall k$ so as to maximise (4.25), $a_k = \Omega_k^{-1}[\sum_{m=1}^{M} u_{km} x_{km}]$. By recalling that $0 \leq a_k \leq 1$, we obtain the following constraints,

$$LT \sum_{m=1}^{M} u_{km} x_{km} \geq 0, \tag{4.29}$$

$$LT \sum_{m=1}^{M} u_{km} x_{km} \leq \Omega_k \quad \forall k \in \{1, 2, .., K\}\,. \tag{4.30}$$

Thus the problem turns into weighted sum rate maximisation with inequality constraints as shown below.

$$\max LT \sum_{k} \sum_{m=1}^{M} \Omega_k^{-1} u_{km} x_{km} \tag{4.31}$$

$$\text{s. t ,} \quad \sum_{k=1}^{K} x_{km} \leq 1 \quad \forall m \in \{1, 2, .., M\} \tag{4.32}$$

$$LT \sum_{m=1}^{M} u_{km} x_{km} \leq \Omega_k \quad \forall k \in \{1, 2, .., K\}\,, \tag{4.33}$$

$$0 \leq x_{km} \leq 1. \tag{4.34}$$

The modified LP in (4.31)-(4.34) can be solved efficiently via primal-dual methods such as LIPSOL [57]. From the relaxation solution, we may develop a further lower bound for solving the admission variables a_k to be binary. For this purpose, let us define an admission list as the subset of users out of the K users who are scheduled resource blocks. The optimal admission list is obtained in a method called relaxation bound as follows.

Relaxation bound

1. Initiate an admission list A consisting of all K users. Solve the LP in (4.31)-(4.34) for the admission list A. Obtain the output bit rates $r_k = LT \sum_c u_{km} x_{km}$.

2. In step t, remove the user with the largest bit rate deficit $\Omega_k - r_k$ from the admission list. This is the new admission list.

3. Solve the LP for the new admission list. Store the admission list and the corresponding bit rates. Go to step $t+1$ until finding an admission list for which all the bit rates $r_k \geq \Omega_k \quad \forall k \in A$.

4. From the stored results in 3) and 4), select the admission list which maximises the primary objective in (4.22).

For most input cases, the linear program gets solved in polynomial time. Still it may not be easy to implement in an online scheduler because of matrix inversion with a large number of variables and sensitivity to the step sizes in the interior point algorithm. We therefore refer to the lower bound as a useful offline algorithm. For faster online techniques, we approach the problem via the Hungarian algorithm [55]. It is noted that optimal solutions result in fractional output values for resource block scheduling to users, which will result in a large downlink control overhead and is also not supported in LTE. Thus binary solutions to scheduling are needed in order to assign an entire resource block to a user as in LTE.

4.3.4 Suboptimal solutions

4.3.4.1 Forward path Assignment Scheme

In this section, we provide fast resource allocation methods which are easy for real-time implementation. Specifically, we look to obtain suboptimal binary integer solutions of x_{km} without performing full enumeration. The binary solutions provide a benefit of manageable overhead in downlink resource map as compared to the real fractional solutions from linear programming. Let us assume that some arbitrary set of resource blocks denoted as $\tilde{M}$ is available for allocation out of the M resource blocks. This means $x_{km} = 0 \quad \forall m \in \tilde{M}, \forall k$. Let us also assume that a user subset $\tilde{K}$ satisfies the constraints in (4.30). Therefore users in $\tilde{K}$ are eligible to be allocated some more subcarriers. We may therefore introduce a resource flow mechanism and allocate L subcarriers to each user in $\tilde{K}$. With this new constraint, the problem becomes

$$\max LT \sum_{k\in\tilde{K}} \sum_{m\in\tilde{M}} \Omega_k^{-1} u_{km} x_{km} \tag{4.35}$$

$$\text{s. t} \quad \sum_{m\in\tilde{M}} x_{km} = 1 \quad \forall k \in \tilde{K} \tag{4.36}$$

$$\sum_{k=1}^{K} x_{km} \leq 1 \quad \forall m \in \tilde{M} \tag{4.37}$$

$$LT \sum_{m\in\tilde{M}'} u_{km} x_{km} \leq \Omega_k \quad \forall k \in \tilde{K} \tag{4.38}$$

$$0 \leq x_{km} \leq 1, \tag{4.39}$$

where we have additionally introduced the constraint in (4.36) and $\tilde{M}'$ is the complementary set of $\tilde{M}$, which are the resources allocated until this point of time. The problem in (4.35)-(4.39) is a linear assignment problem for which the optimal solutions are obtained by the Hungarian algorithm and are known to be binary integers [55]. The benefit from the Hungarian approach is that if x_{km} is required to be binary in the multi-user loading problem, the optimal solutions of this assignment problem can satisfy it. This means that the L subcarriers which are allocated to a user belong to the same resource block. The task is now to update $\tilde{M}$ and $\tilde{K}$ so that (4.37) and (4.38) are not violated. For this purpose we devise a forward path Hungarian heuristic as described below.

1. Start with a user set $\tilde{K}$ which consists of all the K users and resource set $\tilde{M}$ consisting of all the M resource blocks.

2. In step t apply the Hungarian algorithm, to allocate resource blocks from $\tilde{M}$ to users in $\tilde{K}$.

3. Update $\tilde{M}$ by removing the allocated resource blocks so as to satisfy (4.37). Update $\tilde{K}$ based on (4.38) using the latest $\tilde{M}$. Goto step $t+1$.

The steps 2)-3) are repeated until all the users are allocated the guaranteed bit rates or until there are no more resource blocks left.

Illustrative Example : Consider an example as shown in Table 4.2 below. Here U1...U5 are 5 users while RB1..RB5 are 5 resource blocks.

	U1	U2	U3	U4	U5
RB1	12	4	0	10	1
RB2	10	6	4	6	1
RB3	12	4	1	8	2.5
RB4	12	2	5	10	4
RB5	12	2	1	8	2
Demand	12	7	4	20	3

Table 4.2: Example of modulation and coding values in resource blocks for different users. The resource blocks are to be assigned to the users to fulfill a bit rate objective.

Before applying the Hungarian algorithm, the entries of each user is normalised with their respective demands as shown in Table 4.3. Henceforth, we refer to these normalised entries as costs. The

	U1	U2	U3	U4	U5
RB1	1	$\frac{4}{7}$	0	0.5	$\frac{1}{3}$
RB2	$\frac{10}{12}$	$\frac{6}{7}$	1	0.3	$\frac{1}{3}$
RB3	1	$\frac{4}{7}$	0.25	0.4	$\frac{2.5}{3}$
RB4	1	$\frac{2}{7}$	1.25	0.5	$\frac{4}{3}$
RB5	1	$\frac{2}{7}$	0.25	0.4	$\frac{2}{3}$

Table 4.3: Resource block table after normalising the values of each user with the corresponding demand.

	U1	U2	U3	U4	U5
RB1	0	0	0	1×10	0
RB2	0	1×6	0	0	0
RB3	0	0	0	0	1×2.5
RB4	0	0	1×5	0	0
RB5	1×12	0	0	0	0
Demand	12	7	4	20	3

Table 4.4: Output allocation table of resource blocks to users based on the Hungarian assignment scheme

Hungarian algorithm is applied to maximum the sum of the costs in the Table 4.3 and results in an allocation as shown in Table 4.4. With this method it is seen that, only 2 out of the 5 users i.e, U1 and U3 are fully satisfied. The others users are fractionally satisfied. In relation to our original problem in (4.24)-(4.28), a clear shortcoming of the assignment approach is the constraint which we introduced in (4.36)). This constraint has a consequence of enforcing allocations to all the users. So this necessarily does not maximise $\sum a_k$, where a_k is binary. To resolve this issue, we propose an admission control step before the assignment problem. This means that we preset some of the a_k to be zero and maximise (4.24) using the rest of $a_k = 1$. In the following sections, we describe a few techniques to perform the admission control.

4.3.4.2 Random admission lists (*ALG1*)

Admission control can be performed before resource allocation by using statistical estimates. In *ALG1*, we introduce a simple scheme where the number of users for allocation is estimated. Assume that there are M' resource blocks available for allocation. The number of users to be admitted, denoted as k_{est}, is estimated as $k_{est} \leq \frac{M'}{\gamma}$, where γ is the expected number of resource blocks for fulfilling the rate requirements. For example, an estimate of γ can be obtained as the average of the ratios between the demands Ω_k and average MCS levels. The purpose of this estimation is to ensure that overloading does not occur at the scheduler. The k_{est} users to be scheduled, denoted as $A_1 = \{1..k_{est}\}$, are randomly chosen out of $\{1..K\}$. Users in the list A_1 are allocated resource blocks via the forward path Hungarian scheme described in Section 4.3.4.1.

After performing resource block allocation for the list A_1, the estimate k_{est} is updated and the next admission list, A_2, which comprises another k_{est} random users is used. In this way users are admitted both randomly and incrementally for resource block allocation in a TTI. However the scheme performs admission control and resource block allocation as two separate functionalities.

Illustrative example :

Consider the same example as before in Table 4.3. To begin with, we assume that two users $U2$ and $U4$ are randomly admitted for resource block allocation. The following steps then follow.

- Apply the Hungarian algorithm only for the two users U2 and U4. This results in the allocations $x_{22} = 1$ and $x_{41} = 1$ and all the other x_{km} is zero. With this, U2 gets allocated 6 and U4 gets allocated 10 which are less than the demanded values of 7 and 20 respectively.

- Therefore we proceed to a 2^{nd} iteration using the same users and use leftover resource blocks as in Table 4.5. This gives a second resource block to each of the two users. Hungarian algorithm is applied again to the above table to obtain the allocations, $x_{23} = 1$ and $x_{44} = 1$. The allocated rates to U2 and U4 are now 10 and 20 respectively, which fully satisfies the above two users.

	U2	U4
RB3	$\frac{4}{7-6}$	$\frac{8}{20-10}$
RB4	$\frac{2}{7-6}$	$\frac{10}{20-10}$
RB5	$\frac{2}{7-6}$	$\frac{8}{20-10}$

Table 4.5: Cost entries for the 2 users after normalising with the unmet demand after the 1^{st} iteration.

- Now the leftover resource block is RB5 which can be randomly allocated to one of the users U1, U3 or U5 with equal probability of 0.33. It is thus clear that random selection affects the satisfaction probability. However, it can be noted that only U1 can be completely satisfied. U2 and U4 require allocations more than the available. On average, the number of satisfied users is just 2.33.

4.3.4.3 Optimal user set (*ALG2*)

For the joint allocation and admission control problem in (4.24)-(4.28), a search for an optimal admission list is needed. This is because the user rate tuples that are obtained from a combinatorial assignment do not only depend on the number of users, but also on the distribution of the MCS values. Thus user subsets are enumerated and resource block allocation is applied for each enumerated subset. The user subset which maximises the objective in (4.22) is taken as optimal.

Here the Hungarian algorithm can been invoked to reduce the complexity of resource block allocation. Thus overall complexity is dominated by user subset generation. This complexity can be too high for a large number of users, K, mainly because the search for a user subset needs to be performed by varying the cardinality of the subset. Therefore we look at even simpler ways.

4.3.4.4 LBORA scheme (*ALG3*)

To increase the number of satisfied users, in this section we present a novel heuristic scheme called Load Balanced Opportunistic Resource Allocation (LBORA). The idea of this scheme is to integrate admission control in a top-descent scheduling. To this end for low complexity we use a simple selection mechanism which is based on a simple ratio between the rate request and the average MCS level of a user. Selection incurs only a worst case complexity of only $O(K)$. We use a two phase method. In the first phase, sum rate maximisation is performed. For the second phase we adopt an external point descent as in [53]. The idea of an external point descent is that for some input cases the maximum sum rate solution may satisfy the original constraints in (4.18)-(4.21). In such cases, the maximum sum rate solution is the globally optimum solution.

If the constraints in (4.21)-(4.21) are not fulfilled by the sum rate optimal solution, resource blocks are reallocated in the second phase. For the re-allocation, the solution in [53] admits users on the basis of the least penalty to the sum rate. However, in case the constraint set of the multiuser loading problem happens to be infeasible, a least penalty reallocation is not a good descent strategy. The reason is simply because the resource blocks may be reallocated to bit rate demands which are infeasible. To overcome this issue, we use a descent with a sorting technique as follows.

1. Sum rate maximisation: Greedy resource allocation is performed to maximise the sum rate. On a per resource block basis, this is done by

$$k^* = \arg\max_k u_{km}, \quad \forall m. \tag{4.40}$$

   From (4.40) we have obtained initial binary solutions as $x_{k^*m} = 1 \quad \forall m$ and $x_{km} = 0 \quad \forall k \neq k^*, \forall m$.

2. Form sets: Two user sets G_1 and G_2 are formed as

$$G_2 = \left\{ G_2 \quad k \right\} \quad \text{if} \quad \sum_m u_{km} x_{km} < \Omega_k, \tag{4.41}$$

$$G_1 = \left\{ G_1 \quad k \right\} \quad \text{if} \quad \sum_m u_{km} x_{km} - \max_m u_{km} x_{km} \geq \Omega_k. \tag{4.42}$$

3. Resource block pooling: A set of resource blocks V are defined for transfer from G_1 to G_2. A resource block can be transfered from G_1, if none of the user bit rates in that set would go below the guaranteed bit rate.

4. Admission: The main idea of incorporating admission control before resource block allocation is to ensure that only a few users do not over consume the downlink bandwidth at the expense of the others. To this end, we obtain the bandwidth metrics as

$$L_k = \frac{|\sum_{m,k \in G_2} \sum_m u_{km} x_{km} - \Omega_k|}{\sum_{m \in V, k \in G_2} u_{km}} \quad \forall k \in G_2 \tag{4.43}$$

and select a user according to $\tilde{k} = \arg\min_{k \in G_2} L_k$. The metric in (4.43) indicates the amount of additional bandwidth that would be needed by a user k. A user in need of the minimum additional bandwidth is given preference by (4.43). Note that this metric exploits two degrees of freedom : user channel and the user demand. This step is performed in each iteration.

5. Resource block allocation: Re-allocation of resource blocks is made using

$$m^* = \arg\max_{m \in V} u_{\tilde{k}m}. \tag{4.44}$$

Update $x_{\tilde{k}m^*} = 1$, and $x_{km^*} = 0 \quad \forall k \neq \tilde{k}$. The resource block m^* is removed from the set V. Steps 2) to 5) are repeated until: a) all users are given their bit rates, or b) the resource block set V is empty.

Illustrative Example : We now illustrate the LBORA scheme using the same example in Table 4.3 as before.

1. Step 1. Resource blocks are allocated greedily to maximise the sum rate. In our example,

this would result in allocations $x_{1m} = 1 \quad \forall m$ i.e user U1 is given all the resource blocks since this user has the best channels. In order to satisfy the demand of other users, some resource blocks have now to be transferred back from U1.

2. Step 2. We require that the allocation for U1 should not fall below the demand of 12. As a result, RB5 is not considered in the pool of resource blocks available for transfer (in order to satisfy the demand of U1). Thus currently RB1, RB2, RB3 and RB4 are the pool of available resource blocks. The bandwidth likelihoods for users U2, U3, U4 and U5 are computed to be $\frac{7}{17}, \frac{4}{10}, \frac{20}{34}, \frac{3}{8.5}$ respectively using (4.43). The minimum requirement is from user U5 and thus U5 is prioritised.

3. Step 3. The best resource block of U5 is RB4, and so we allocate $x_{54} = 1$. RB4 is now removed from the pool of resource blocks. U1 is now satisfied. The bandwidth likelihood metrics are again computed for U2, U3 and U4 as $\frac{7}{14}, \frac{4}{5}, \frac{20}{24}$ respectively.

4. Step 4. Among the three leftover users, U2 requires the least bandwidth and is thus prioritised. The best resource block for U2 is RB2 and thus set $x_{22} = 1$. RB2 is now removed from the pool of resource blocks. The demand of U2 has now reduced to 7-6=1 and the available resource is 14-6=8. The bandwidth likelihood for U2, U3 and U4 are $\frac{1}{8}, \frac{4}{1}, \frac{20}{18}$ respectively. U2 again requires the least bandwidth, is therefore prioritised and allocated its best resource block RB3, i.e $x_{23} = 1$.

5. Step 5. The last remaining resource block is RB1. The bandwidths likelihoods are calculated for U3 and U4, which are $\frac{4}{0}, \frac{20}{10}$ respectively. Accordingly U4 is selected and allocated the resource block RB1 . Based on these allocations, U1, U2 and U5 have been allocated a total of 12, 10 and 4 respectively and are fully satisfied. From this example we see that user admissions based on user ordering clearly improves the user satisfaction to 3 as compared to 2.33 from random admissions.

4.3.4.5 Extended LBORA scheme (*ALG4*)

ALG3 obtains a solution for resource block allocation through a sorting technique. The sorting scheme is however not combinatorial. A combinatorial scheme (e.g, Hungarian algorithm) may provide better rate allocations by exploiting the distribution of MCS levels of a user. However, it was seen that to perform a combinatorial assignment, we require a user set search so as to incorporate admission control. As a means to reduce complexity, instead of a full enumerated search as in *ALG2*, we may start the search from the admission list solution of *ALG3*. Thus we call this scheme, the extended LBORA scheme.

Thus the subset of users who are allocated the bit rates in *ALG3* can be regarded as the initial admission list. The admission list is updated after scanning through possible additions to the list for maximising the objective in (4.22). We therefore need to obtain a feasibility certificate for possible additions. To this end, a valid allocation for multi-user bit loading of an admission list acts as its feasibility certificate.

Suppose A is the initial admission list i.e, the set of users who have been allocated their required rates in *ALG3* and A' is the complementary set of A. Now the problem is to find the best incremental subset Z^* as

$$Z^* = \arg\max_{Z \subset A'} |Z| \tag{4.45}$$

$$\text{s.t} \quad LT \sum_{m=1}^{M} u_{km} x^*_{km} \geq \Omega_k \quad \forall k \in \{Z, A\} \tag{4.46}$$

$$x^*_{km} \in \{0, 1\}, \tag{4.47}$$

where x^*_{km} is the optimal integer solution for resource block allocation. The complexity of a brute force resource block search is upper bounded by $(|A| + |Z^*|)^M$.

For low complexity, the Hungarian algorithm can be used to obtain a feasibility certificate for the user set. Even though this approach is sub optimal for verifying feasibility, it can be noted that the max operation in (4.45) captures the maximum over all the suboptimal values and increases Z^*.

There may be cases where the feasibility certificate needs improving. Notably, assume that there are only $|A''|$ with feasible bit rates out of the $|A'|$ users to be scanned. By starting from the admission list solution of *ALG3* we have ignored other user subsets of the same cardinality $K - |A'|$. Some of these subsets may have also been feasible. To note the difference, the search in *ALG2* enumerated all the possible subsets but *ALG4* starts with only one subset. Therefore, if $|A''|$ is very small, the max operation in (4.45) becomes redundant.

Thus, we may need some more robust resource block allocation solutions. For this purpose, we can generalise the constraint in (4.36)-(4.39) as

$$\sum_{m \in \tilde{M}} x_{km} \leq 1 \quad \forall k \in \tilde{K}, \tag{4.48}$$

$$\sum_{k \in \tilde{K}} \sum_{m \in \tilde{M}} x_{km} = y, \tag{4.49}$$

$$0 \leq x_{km} \leq 1, \tag{4.50}$$

where y is a real integer greater than zero and less than K'. $\tilde{M}$ and $\tilde{K}$ are the pool of leftover resource blocks and users respectively at any stage of iteration.

The assignment problem (4.48)-(4.50) is known as the k-cardinality assignment problem (in our case renamed as y-cardinality assignment problem). The y-cardinality assignment basically assigns y resource blocks at a time in a $\tilde{K} \times \tilde{M}$ assignment problem. This problem can be solved quickly using the Hungarian algorithm by making a simple transformation and the optimal solutions are binary integer [104]. For the multi-user loading problem, y-cardinal sub-problems are implemented on a forward path heuristic similar to Section 4.3.4.1. At any stage of the scheme, the path depth $y = min(\tilde{K}, \tilde{M}, \tilde{y})$, where $\tilde{y}$ is a preset value. We term this a forward path y-cardinality scheme. Importantly, for each value of $\tilde{y}$, a different resource block allocation solution and feasibility certificate result from the forward path y-cardinality scheme. We can thus simply choose the best feasibility certificate from all the values of $\tilde{y}$. An implementation of the forward path scheme is described below to maximise (4.48)-(4.50).

1. Initialise iteration as $t = 0$. Run *ALG3* to obtain $x_{km} \quad \forall k, \forall m$. Find the list of users with $a_k = 1$ according to (4.22) for all k. The computed list of users is stored as the initial admission list A. Calculate $g_0 = \sum_k a_k$.

2. Initialise a set A' which is the complementary set to the admission list A.

3. Let h_n be the n^{th} user in A'. Obtain $M_n = \left\{ \begin{array}{cc} A_0 & h_n \end{array} \right\}, \quad \forall n$. Implement the forward path y-cardinality scheme for all $\tilde{y}$ from $\tilde{y} = 1$ to $g_0 + g_t'$ and apply it on M_n for all n. Increment $t = t + 1$. For each $\tilde{y}$ and n, calculate $g_t(\tilde{y}, n) = \sum_k a_k$ according to (4.22).

4. Update the values of g_t as $\max_{\tilde{y}} \max_n g_t(y, n)$. Let the optimum value of n be n^*. If $g_t \geq g_{t-1}$ increment $g_t' = g_{t-1}' + 1$, set $A = M_{n^*}$ and return to step 2). Otherwise exit with g_{t-1} and A as the output.

Illustrative Example : The working of this scheme is illustrated using the same example.

1. Step A. We first run the the LBORA scheme. Based on this, we already know that the data rate demand for U1, U2 and U5 are feasible. Thus the first admission set A comprises of these three users.

2. Step 2. The complementary set A' consists of users U3 and U4. Out of these two users, U3 requires the least amount of resources while U4 requires substantially more resources and may be infeasible. At this stage we have no prior knowledge about the feasibilities of the two users. We add U3 to the admission set to test for feasibility. We now need to obtain the feasibility certificate of (U1, U2, U3, U5) with high reliability. For this purpose, we use the y-cardinality assignment scheme.

3. Step 3. The path depth $\tilde{y}$ in the y-cardinality scheme is varied from 1 to 4. For instance, consider a path length $\tilde{y} = 2$. This means 2 RBs are allocated at a time for 2 out of the 4 users. Accordingly, RB5 is given to U1 and RB4 to U5 (with a cost metric $1 + \frac{4}{3}$) from Table 4.3.

4. Step 4. U1 and U5 have now been given 12 and 4, and therefore both are satisfied. Subsequently, the two users U1 and U5 are removed from the user list, while RB5 and RB4 are removed from the resource block list. The leftover users are U2 and U3.

5. Step 5. The best cost metric for the second block of allocation is $1 + \frac{4}{7}$ which corresponds to allocating RB1 to U2 and RB2 to U3. These allocations so far satisfy users U1, U3 and

Method	Admission control	Output allocation variables x_{km}	Reference
Linear program	No	Continuous	(4.31)-(4.34)
Assignment	No	Binary	(4.35)- (4.39)
LBORA	Yes	Binary	(4.40)-(4.44)
Extended LBORA	Yes	Binary	(4.45)-(4.47)

Table 4.6: Table of reference for methods described for guaranteed resource allocation

U5. The leftover resource block is RB3 and pending user (dissatisfied user) is U2. The path length for this allocation is set to min (1,1,2) because there is only one resource block and user left. RB3 is allocated to the remaining user U2 upon which U2 is fully satisfied. Thus we have obtained $x_{54} = 1, x_{15} = 1, x_{32} = 1, x_{21} = 1, x_{23} = 1$, based on which 4 users are satisfied.

6. Termination condition : Now we apply the same steps above to the user set (U1, U2, U4, U5). It can be seen that this set is not feasible. Thus we retain the set (U1,U2,U3,U5) as the optimal. It can further we easily seen that for this example, this optimal set matches with the output of a brute force search.

A tabular summary of the methods we have discussed for guaranteed bit rate allocation is shown in Table 4.6.

4.3.5 Simulation and Results

4.3.5.1 Complexity analysis

Table 4.7 presents the complexity of scheduling techniques on the basis of sort complexity. In the case of PF, in any m^{th} iteration, there are $M - m + 1$ resource blocks in the pool. For each resource block K user comparisons are made and then a comparison of $M - m + 1$ resource blocks is done. As a baseline, the complexity of finding the maximum in an array of n elements is defined as $O(n)$, and called the selection complexity. Other complexities are derived on this basis. For example, the complexity of sorting n elements in an array becomes $O(n^2)$.

The LBORA scheme that we propose in Section 4.3.4.4 starts with a sum rate maximisation step.

The initial sum rate maximisation computation has a complexity of $MO(K)$, because K user comparisons are made sequentially for M resource blocks. Further to this, the scheme computes the resource block allocation iteratively. The worst case iterative complexity occurs when all the M resource blocks have to be re-allocated. Using this upper limit, we break down the complexity of each step as follows. In iteration m, pre-selection of user based on bandwidth feasibility likelihoods in the admission control step has a worst case complexity of $O(K)$. This has to be done for a maximum of M iterations and thus a total of $MO(K)$. There are $M - m + 1$ resource blocks in the pool and thus the complexity of selection is $O(M - m + 1)$ in the m^{th} iteration. Note that a resource block is removed from the pool after each iteration. Furthermore, resource block comparisons are performed only for the pre-selected user in each of the M iterations. Put together, the total complexity is written as a sum of the complexities of all steps as shown in Table 4.7.

In the case of extended LBORA scheme, possible users are added to the admission list after scanning for feasibility certificate. In iteration k ($k > 0$), let there be $K - k + 1$ possible users who can be added to the list. In the worst case, the feasibility certificate has to obtained for each of the candidate users (because the last searched user may happen to be the feasible one). Thus a multiplicative factor of $K - k + 1$ is incurred in complexity in each iteration k.

In obtaining the feasibility certificate for k users and given M resource blocks, the y cardinality algorithm incurs an expected complexity of $0(y(k + M - y)^2)$ as shown in [104]. It can be noted that the complexity varies with the value of y. This is evident because for $y = 1$, it is simply breaks down to a selection algorithm.

In extended LBORA scheme, the y-cardinality algorithm is repeated for various path depths y which ranges from 1 to $K - k + 1$. For a path depth y and with M resource blocks, the forward path algorithm repeats the process approximately $\frac{M}{y}$ times.

4.3.5.2 Study methodology

Numerical Results We illustrated the working of the proposed schemes using a numerical running example. For more robust evaluation of expected performance effectiveness, we use a simulation

Method	Time Complexity
Full binary search	K^M
Proportional fairness	$\sum_{m=1}^{M} O(M-m+1) + (M-m+1)O(K)$
LBORA	$2MO(K) + \sum_{m=1}^{M} O(M-m+1)$
Extended LBORA	$\sum_{k=1}^{K}(K-k+1)\sum_{y=1}^{k}\frac{M}{y}O(y(k+M-y)^2)$

Table 4.7: Complexity of the scheduling techniques derived from selection complexity

based study. To evaluate the performance, we assign randomised cell positions to users in a channel realisation and generate 1000 such channel realisations from the stored macrocell measurements in 2.6 GHz. The number of served users is obtained as the result of a desired algorithm in each channel realisation. For the plots, the arithmetic mean of the number of served users is calculated over the 1000 channel realisations. The dimensions of the measurement location are approximately 300 m $\times$ 160 m as in [116]. The measurements collected a total of 3500 samples from both line of sight and non line of sight conditions from a base station with a transmit power of 43 dBm. These measurements are the same as that used for numerical results earlier in Section 4.2. The radio interface is a closed-loop FDD 4 $\times$ 2 MIMO system which was implemented using 3GPP working assumptions in [4]. Thus two spatial streams are available in each subcarrier of a resource block. A resource block consists of 25 sub-carriers and 7 long OFDM symbols.

Study scenario.

For simulation study, the channel values from the measurement samples were normalised so that the average received power is unity. The channels were multiplied by an average signal to noise ratio (SNR) of 15 dB. We remark that an average SNR value of 42.2 dB was realised in multiple measurement tests in an interfere-free scenario in that cell. However, a lower SNR is assumed for the simulations to conservatively account for the possibility of larger cell sizes. The received SNR is estimated for the two spatial streams in each resource block at the user equipment. From the computed SNR, the MCS level per spatial stream is obtained from the mapping table in Table 4.1, [109]. The mapping table defines 26 modulation and coding schemes (MCS) for a target packet error rate of 10^{-2}. The lowest MCS level is of spectral efficiency 0.5 b/s/Hz and the highest is 5.5 b/s/Hz per spatial stream.

We assume that there are K users at different positions in the measurement track and thus experience different channel frequency responses. They have to be scheduled in a transmit time interval (TTI) and allocated a specified bit rate R. As an objective, we apply admission control on the K users and fulfill the bit rates of the users who qualify for admission. To this end, resource blocks are allocated based on the MCS levels. As a performance measure, we calculate the number of fulfilled users out of the K users assuming that the users' channels are static. A statistical average for the macrocell is computed by randomising the user positions which in effect averages over the sets of measurement channels. By varying the values of K and R, different network congestion levels are obtained. In a real-time system, the number of users K can change over different TTIs depending on user and packet arrivals. In our model, we capture the probability of providing their bit rates as a function of K. The following schemes are simulated: a) *ALG1* (random admission (RA) with Hungarian) b) *ALG3* (user ordering), c) *ALG4* (incremental user set search) and d) relaxation bound in Section 4.3.3.1. *ALG2* is not implemented because of high complexity. We refer to *ALG3* and *ALG4* as fast search schemes. For baseline comparison, we simulate a round robin allocation with the random admission scheme in *ALG1*. For *ALG1*, the value of γ is set to 6, i.e $k_{est} = 8$ for $M = 48$.

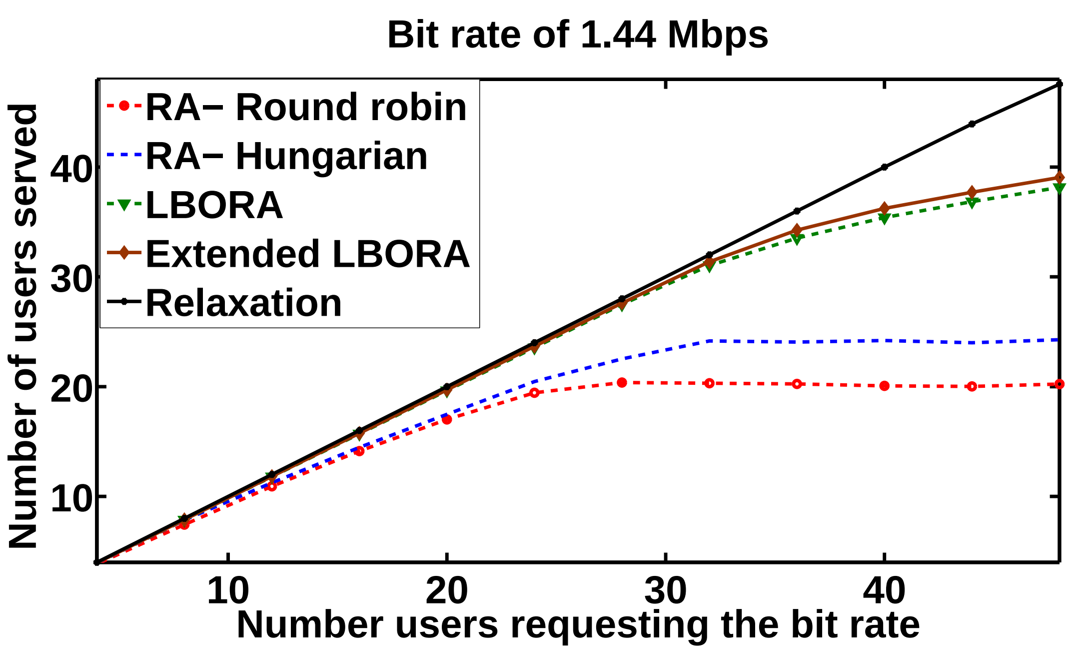


Figure 4.6: Performance of LBORA and other scheduling schemes for guaranteed bit rates of 1.45 Mbps (low bit rate request)

Effect of number of users. For this study, the number of users K in a TTI is varied from 4 to 48 in steps of 4. We consider a maximum of 48 users so that each user can be assigned at least one resource block out of the total 48 resource blocks. As the number of users increase, it means that the user density within the macrocell increases. In Figure 4.6, the users are assumed to require 720 bits to be loaded in a 0.5 ms transmit time slot, which would correspond to 1.44 Mbps served rate. In a simple case, if each user is allocated one resource block and loaded 5 b/s/Hz MCS in a spatial stream, all the users can be satisfied. In real-time, the realised bit rate at user end will be lower than 1.44 Mbps based on the serving interval, i.e, the interval between two servings of a user.

X- axis shows the total number of users who request this bit rate. Y- axis shows the number of users who are loaded with the required bit rate after applying admission control and resource block allocation. The scaling law between the two axes is termed the network scaling ratio.

For this relatively low bit rate, random admission using $ALG1$ (red-circle and blue-dotted lines) is seen to be a good enough strategy and even for an arbitrarily large number of users. There are two notable aspects of the algorithm: random admission and increments. Random admissions can handle up to 20 user requests. Some of these users can be at a relatively large distance to the base station because the user positions are random. At these large distances, users benefit from the different MCS levels on resource blocks. Both round robin and Hungarian methods are able to perform bit loading by making use of the MCS levels. Increments prevent overloading when there are too many users. An efficient allocation scheme, such as the Hungarian method, shows a 20% gain as compared to the round robin scheme. Integer solutions for joint admission and allocation (fast search schemes) outperform the random admission schemes. The gain that is achieved by the LBORA scheme as compared to Hungarian scheme is significant 58% (increase from 24 to 38 users). The extended LBORA scheme adds one more user to 39 users. The heuristic LBORA schemes perform very close to the linear programming solution for up to 28 contending users, after which the gap widens. For 48 contending users, the optimal relaxed solution can support all the 48 users.

Effect of bit rate. For this study, a higher loading of 5.8 Mbps is assumed for each users. Results are shown in Figure 4.7. This data rate rate is considered to be high because bit loading cannot

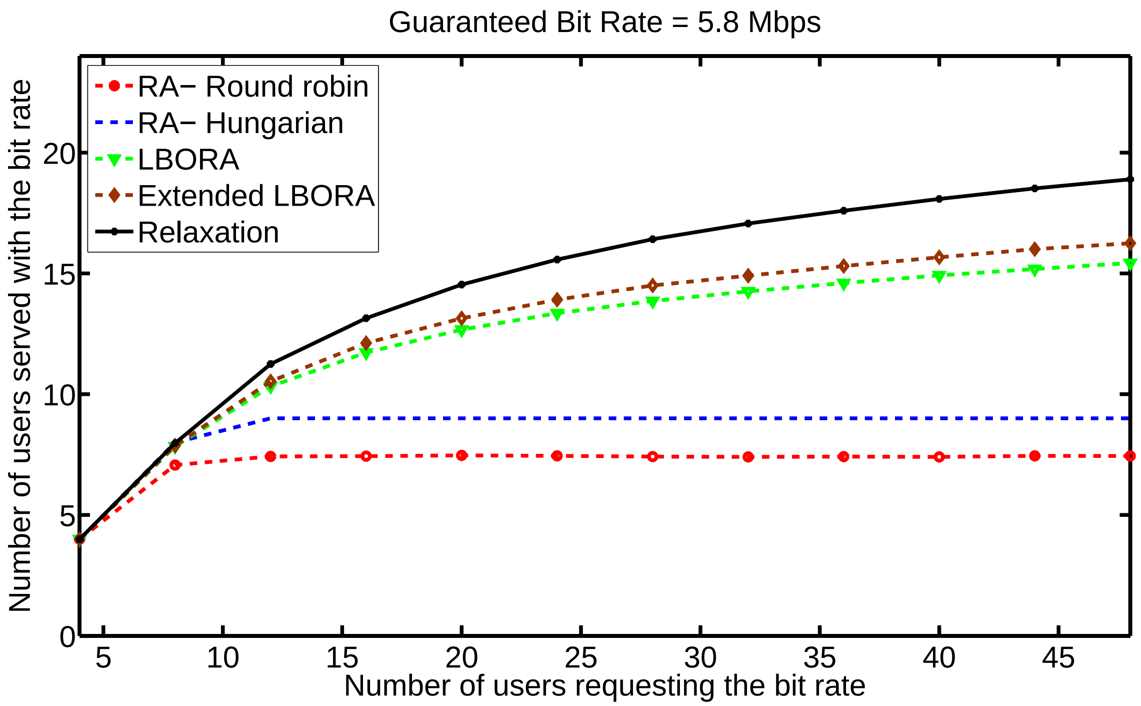


Figure 4.7: Performance of LBORA and other scheduling schemes for guaranteed bit rates of 5.8 Mbps (high bit rate request)

be done by allocating just one resource block to a user. In a practical network, contents may be available in different qualities. So we have simulated the schemes by assuming a low and a high bit rate. Similar to the low bit rate case, the Hungarian method is again more efficient, by 20%, than the round robin assignment. The LBORA scheme can support 16 users as compared to 9 with Hungarian with random admissions. This represents a 78% gain. We also observe that these fast allocation schemes (LBORA and extended LBORA) are even more effective in high bit rates, up from 58% in Figure 4.6 to 78% for the higher bit rate in Figure 4.7. The key idea in LBORA scheme is to integrate admission control in resource block allocation. Whenever the network load is high, either because of high number of users or high bit rate demand, admission control is thus found to be beneficial. This is logical because we prevent a few users from overconsuming the bandwidth. In effect, by denying admission to users with high bandwidth requirement we increase the chances of fulfilling other users. It is also seen that the extended LBORA can admit at least one more user in both the scenarios using some more search complexity. In the simulations, the relaxation bound proves to be a useful upper bound for the fast search schemes. With relaxation, 18 users can be supported when 48 users ask for service. The continous relaxation of allocation variables efficiently exploits the different MCS levels in resource blocks.

In both results, Figure 4.6 and Figure 4.7, it is observed that the performance of simple LBORA scheme is close to the upper bound achieved by linear programming based optimal solutions. In the case of 1.44 Mbps, optimal solution performs 26% better than LBORA (compare 48 users to 38 users). Furthermore, for 1.44 Mbps case, the performance of LBORA matches the upper bound for upto 28 users as compared to just 12 users being the matching point for the Hungarian scheme. In the case of 5.8 Mbps, optimal solution performs 12.5% better than LBORA (compare 18 users to 16 users). These observations are quite remarkable for the following reasons : a) LBORA scheme is quite easy to implement, b) the running-time complecity of LBORA is much lower than linear programming, c) the resource block allocation solutions of LBORA are binary and thus would incur lower downlink control overhead as compared to optimal solutions.

4.4 Conclusion

Resource allocation can be classified into non-guaranteed or guaranteed bit rate allocations. This type of resource allocation assigns resource blocks to users with some fairness but without an obligation to fulfill their bit rate demands. They can be applied for non-guaranteed bit rate service mode. We further show two different types of non-guaranteed allocations : rate proportional and utility proportional. Numerical results obtained from measurements show performance within 5% of optimum and 71% gain over roundrobin scheduler.

Guaranteed bit rate (GBR) allocations need to fulfill the bit rate demand of users. In this case, a new performance metric to maximise the number of GBR users is proposed. We also show dynamic resource block allocations that can achieve this objective. The novelty is to integrate admission control in the process of resource block allocation, which we term joint admission control and resource allocation. We show that the optimal allocation to the joint problem can be obtained via solving a linear program. However linear programs can be quite difficult to solve in real-time and within the channel coherence time. We thus present a fast two step heuristic algorithm called Load Balanced Opportunistic Resource Allocation (LBORA). The algorithm performs a greedy allocation first and then a descent with admission control. Simple sorting operations are used which can be easily implemented. Numerical results obtained from measurements show an increase from 24 to 38 users at a data rate of 1.45 Mbps. An even larger performance gain of 71% from 9 users to 16 users is observed for a higher 5.8 Mbps. This is obvious because the algorithm prevents overloading of the system at higher data rates by only few users. The performance of LBORA is within 26% of the optimium.

This chapter has shown several resource allocation techniques and presented results in an outdoor to outdoor macro-cell environment. In reality, many users may be located indoors, where the fading loss is high. Outdoor to indoor coverage can be a challenge because a maximum transmission power limit exists at the macro-base station. Thus means to extend coverage are needed especially when the link PER is required to be very low (10^{-6}) and when only limited MCS feedback can be sent through the uplink. In the next chapter, we focus on this issue and propose relaying as a solution to improve coverage to indoor users.

Chapter 5

Relaying in 4G

5.1 Relaying in cellular networks

In the last decade, there has been a revived interest in exploiting wireless relaying as an additional source of diversity. In cellular networks, the use of relays have been in the form of fixed repeaters; a device which boosts the signal and transmits again. Repeaters are usually installed on a mast by a service provider and transmission parameters are set to fulfill the network needs and standardisation requirements. The main idea is to increase the signal power at the end user, which is perceived as coverage improvement. Over the last few years, the cellular subscriber base has grown manyfold. In view of the recent improvements, one might wonder how repeaters and relays are relevant to OFDM cells.

Recent LTE studies based on channel models have shown the presence of so-called coverage holes in macrocells [11]. These are areas in the cell which receive markedly low signal power. Coverage holes with degraded throughput have been observed in recent measurements conducted at 2.6 GHz [116] which agrees with the simulation studies in [11]. Low signal power can be caused by multipath fading, a well recognised problem in wireless channels. The problem further accentuates if the mean path loss is high. It is well known that the mean path loss is proportional to the carrier frequency and the path loss exponent. For outdoor to indoor communication, there could be further attentuation losses arising from building penetration. A compilation of measurement studies on concrete building penetration losses for carrier frequencies 912, 1920 and 5990 MHz has been done in [22], which provide values of 7.7,11.6 and 16.1 dB respectively. The above study notes that building penetration losses because of concrete wall increases as the carrier frequency increases. Thus building (concrete) penetration loss and modern window coatings originally designed for thermal insulation can potentially generate high attentuation at 2.6 GHz. Relays can be used to mitigate some of these issues in OFDM cells.

A step in this direction has been initiated by ITU by setting an average cell spectral efficiency requirement of 3 bits/s/Hz/cell indoor users [13] on fourth generation systems (4G). LTE-Advanced [12] has been working on indoor coverage solutions to meet the target. Deployment of more advanced relaying by data regeneration, also known as decode and forward (DF) is under discussion. We have provided a literature survey of these relaying techniques in Chapter 3.

System level investigations with DF relays have been performed for OFDM cells in [11] where improvement in cellular network performance is shown. In the indoor case, relaying has been observed to improve the 5 percentile outage by 3 dB.

Our main contribution in this chapter is to propose a DF relay scheme and perform indoor field trials with multi-antennas at 2.6 GHz and 20 MHz bandwidth. Two sets of measurements are performed with the MIMO-OFDM air interface using channel quality feedback. A direct outdoor to indoor measurement from a nearest base station at a distance of approximately 500 m. The second is a set of DF relay measurements from an indoor relay. The measurements show the effectiveness of relaying in a typical indoor scenario.

Furthermore, performance assessment of the achievable rate is presented using adaptive modulation and coding. We classify the different types of feedback type as coarse, fine or extended. Coarse feedback represents the channel quality of the full bandwidth. Fine feedback represents the channel quality of smaller bandwidths (resource block based feedback). Extended feedback adds precoding feedback to the fine scheme.

In the proposed relay concept, an indoor relay is used in DF operation with very low packet error

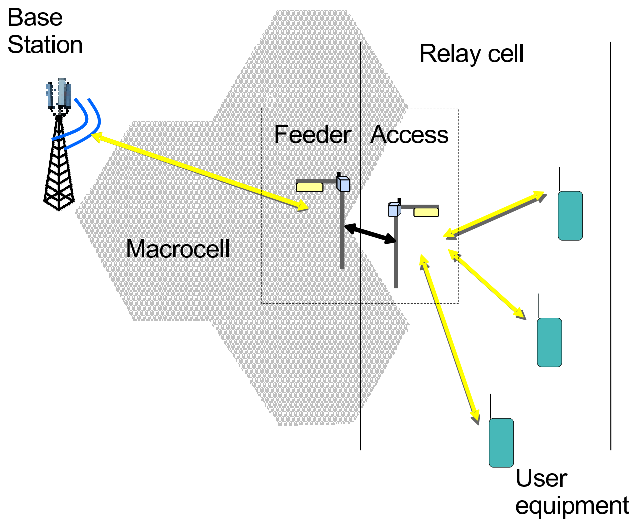


Figure 5.1: A cellular three terminal relay set-up

rate. Because of insulation of the building to outdoors, the relay generated interference is isolated from the adjacent macro-cells. A DF relay exploits the situation and reuses the full bandwidth for forwarding to the end user.

The measurements look at the basic but paramount question: will relaying improve coverage and data rate to the indoor user ?

5.2 Relaying scenario

An indoor wireless relay configuration is shown in Figure 5.1. The macro base station serves the users in its cell, which has dimensions in the order of hundreds of meters. Most locations in the macrocell will receive sufficient signal strength and realise data rates above a satisfactory level. These locations comprise the coverage region of the base station. However, there might be locations in the cell where the data rate is not satisfactory. Wireless relaying is one way of improving the data rate. From an operator perspective, the main advantage of wireless relaying is cost-effectiveness because additional backhauling is not required. For example, installation of expensive cables or fiber optic links is not necessary. A user is given handover to the relay if the received signal power from the relay is higher than that from the base station. In a handover situation, the data transfer between the relay and macro base station is done on the same air interface as the rest of the cell. The relay re-transmits the incoming signal after regeneration of the data, termed as a DF operation. The coverage area of a relay is an order of magnitude smaller than a macrocell and is called a relay cell.

In a typical installation, the relay can be placed at a convenient location indoors by making use of the existing infrastructure, a deployment which is similar to a Wifi access point. However, it is envisioned to be the task of a service provider. The relay unit consists of two parts: a feeder unit which connects to the macro base station and an access unit which connects to the user equipment. Both the units use 3GPP LTE-like MIMO-OFDM air interface of our test-bed in Section 2.2.

5.3 DF Relaying Scheme for LTE

The isolation between the relay feeder and access for 1 m separation at the measurement site and with additional shielding was only about 20 dB. In case of low isolation, orthogonalisation between the two relay units is needed. A two phased FDD strategy shown in Figure 5.2 was employed for this purpose. We describe the downlink.

In phase 1, the relay feeder unit functions as a receiver terminal and receives data from the macro base station for some OFDM time slots of a radio frame. Notably, the cyclic prefix is removed and time frequency synchronisation is achieved with the base station. Demodulation and full decoding of IP packets are performed with a Viterbi decoder. The IP packets are then re-numbered and sent via an Ethernet cable to the relay access unit to complete the phase 2. In phase 2, after re-encoding the information bits, the relay access unit applies a power level of 23 dBm with uniform power spectral density to the indoor LTE-like air interface. There is no necessity for a guard period between the two phases operating in FDD mode, as noted in [114]. The link adaptation for the relay to user equipment is run independent of the base station to relay link. A standard cyclic prefix length of 4.7 μs (7 OFDM symbols per subframe) and subcarrier spacing of 15 kHz are used with a total of 1200 used subcarriers. Duplexing between uplink and downlink is made using frequency division duplexing (FDD) for the two hops: base station to relay and relay to user. Therefore

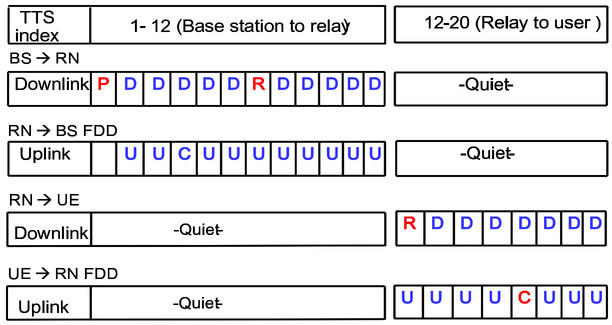


Figure 5.2: Timing diagram of decode and forward FDD relaying. D-downlink data, U- uplink data, R- radio resource management, P- AB/TA preambles, C-channel quality indicator, TTS - transmit time slot.

overall there are four point to point links: two downlinks, from macro base station to relay and relay to user equipment, and, vice versa two uplinks. Data transmission in the two uplinks are performed with single carrier FDMA (SC-FDMA). The working of four duplex links are shown in the same Figure 5.2 for a time duration of 10 ms with 20 time slots. The figure shows uplink feedback and downlink control signaling in a radioframe. AB and TA are the downlink random access and uplink timing advance preambles respectively for synchronisation. In the rest of the chapter, we focus on the data rate of the downlink.

5.3.1 Derivation of efficiency

In practice, a base station may not be able to dedicate the full transmission bandwidth for the phase 1 (feeder link) to one relay. The reason is because there are other relays and/or users in the cell who need to be served. Therefore the phase 1 bandwidth need not be equal to the phase 2 bandwidth. In fact, we find that this bandwidth asymmetry can be exploited for network efficiency because it helps to overcome the relay half-duplex loss that is realised to be significant in [71]. The downlink data transfer in N time slots per relay cell is realised as

$$A_r = \min\,[T_1 \sum_{f_1 \in B_1} d_1(f_1), T_2 \sum_{f \in B} d_2(f)] \tag{5.1}$$

$$\text{s.t} \quad T_1 + T_2 = N. \tag{5.2}$$

Here T_1 and T_2 are the durations of phase 1 and phase 2. B_1 is the phase 1 frequency band, i.e the set of resource blocks for phase 1. B is the band that is used for phase 2 and is assumed to be the full allocated bandwidth. d_1 and d_2 are the data rates supportable for resource block f_1 and f respectively. Upon inspecting (5.1), it can be deduced that the optimising T_2 as a function of B_1 is

$$T_2^*(|B_1|) = \frac{N \sum_{f_1 \in B_1} d_1(f_1)}{\sum_{f_1 \in B_1} d_1(f_1) + \sum_{f \in B} d_2(f)} \tag{5.3}$$

Combining the results in (5.1) - (5.3), the spectral efficiency of a relay cell w.r.t bandwidth $|B_1|$, denoted C_r becomes

$$C_r = \frac{A_r}{|B_1|} = \frac{xg}{g\Delta + x}, \tag{5.4}$$

where we have replaced $\sum_{f_1 \in B_1} d_1(f_1)$ with $g|B_1|$, $\sum_{f \in B} d_2(f)$ with $x|B|$ and $\frac{|B_1|}{|B|} = \Delta$ in (5.3). g and x can be understood to be the average spectral efficiency on the feeder link and access link respectively. From (5.4), it is observable that spectral efficiency C_r is a monotonically decreasing function of the ratio Δ, i.e, if $|B|$ is a constant, then C_r monotonically decreases with input bandwidth $|B_1|$. This result suggests that a network which reduces the feeder bandwidth to the relay can increase its spectral efficiency. Thus intuitively making many relay cells active at the same time using spatial reuse can improve the network efficiency.

The phase 1 relay bandwidth cannot vanish to zero even though it benefits the network. This is in order to satisfy a second measure known as user spectral efficiency U_r. This efficiency is given by the number of correctly received bits in the time duration N over a bandwidth $|B|$ and described by the relation

$$\frac{1}{U_r} = \frac{A_r}{|B|} = \frac{1}{x|B|} + \frac{1}{g|B_1|}. \tag{5.5}$$

Comparing (5.4) and (5.5), one can perceive a trade-off between the two measures for a relay cell with a fixed $|B|$. While C_r decreases against $|B_1|$, U_r increases with $|B_1|$. However both the measures increase with x, which is the mean spectral efficiency of phase 2. An operator policy may be to enforce a minimum user efficiency and maximise the network efficiency on top of this

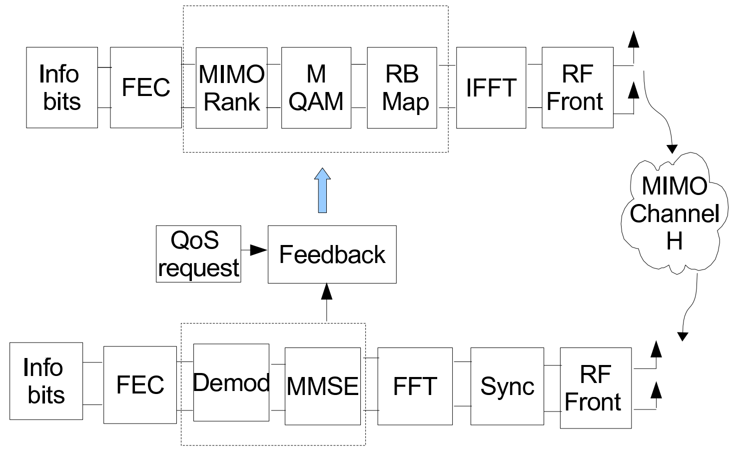


Figure 5.3: Functional blocks of OFDM air interface transceiver

constraint or vice versa. One easy way of achieving this is to set the bandwidth $|B_1|$ to a minimum level $|B_{min}|$, to give a minimum data rate to each relay.

5.4 Classification of feedback schemes from the user to relay

We now look at the transceiver architecture more closely as shown in Figure 5.3 for the downlink. This architecture is used by both the base station and the relay and was applied in the test-bed for taking the coverage measurements which will be shown later. Typically, the packet error rate (PER) target is a part of the quality of service (Qos) request and depends on the transport layer protocol and the traffic class. In most cases, the link requirements of transport control protocol (TCP) is considered to be stringent. The main issue is that it demands a very low PER (10^{-6}) and ordered delivery in order to perceive a loss-less channel [106]. This is in turn can adversely impact the availability of coverage.

For reasons of simplicity, the very low link PER is targeted within the first transmission without taking into account re-transmissions using hybrid automatic repeat request (H-ARQ). However, in the event of any packet errors in first transmission, the low round trip time of 8 ms from relay to user in LTE can still be utilised to fully correct those packet errors using HARQ. The role of adaptive modulation here is to maximise the spectral efficiency with a guaranteed uncoded bit error rate (BER) which in turn can be used to control the PER. External to the adaptive modulation, a forward error correction (FEC) code (known as channel code) is used to obtain the target PER from the uncoded BER. The low first transmission PER requirement may limit the downlink data rate at the physical layer, and as a result we call this TCP coverage. The transmitted downlink bit rate to a user in bps can be calculated as

$$C = \frac{N_b * N_d * \beta}{T_s} \quad \text{bps}, \tag{5.6}$$

where N_b is number of coded bits that is loaded per OFDM symbol (dependent on the SNR), N_d is total number of OFDM symbols carrying data, T_s is an OFDM symbol period and β (0.95) is a scaling factor for downlink control signaling overhead.

In what follows, we describe methods for adaptive modulation and coding in the downlink based on the amount of feedback. Three feedback schemes classified as coarse, fine and extended are described based on the amount of feedback that is sent on the uplink. In the presentation of results, the achievable gain in data rate with a multi-antenna relay is shown for each feedback scheme.

5.4.1 Coarse feedback

In the coarse feedback scheme, the least amount of feedback information is sent to the base station. A practical motivation is that feedback of less than 20 coded bits can be sent via the existing uplink control channel called PUCCH. Suppose there are two transmit and receive antennas and we use a scheme only to select the MIMO rank [1] and modulation order. For this case, 2 feedback bits per spatial stream can convey the selection of modulation order, rank and transmit antenna selection (transmit diversity) with a modulation set of : a) M=0 (no transmission), b) M=4 (QPSK), c) M=16 (QAM) and d) M=64 (QAM). These feedback bits can be further protected to less than 20 bits with channel coding for reliable transmission. Still, a difficult task is to compute a single modulation order that is supportable on all the frequency subcarriers. The mechanism is computationally intense because to predict the PER, one has to take into account channel coding,

[1] The term MIMO rank is often used to denote the number of spatial streams

channel fading distribution and as well as the modulation order [111]. But as shown in Figure 5.3, we bypass this computational complexity by only tuning the uncoded BER to a required level and then achieve the target PER. The tunable uncoded BER controls the adaptive modulation. To quantify the switching points of the adaptive modulation, we use the uncoded BER bounds as derived in [108] for AWGN channel as a function of SNR. A fixed channel coding is used to obtain the target PER from the uncoded BER. The relation between uncoded BER and the PER can be made using the mapping in [110].

Adjacent subcarriers are grouped into the so-called resource blocks. The channel matrix of the center subcarrier in a resource block is used for MMSE inversion. The SNR per spatial stream is then calculated and applied to the whole resource block. Let $\frac{S}{N}(s,k)$ be the SNR of stream s in resource block k and M_s the modulation order of stream s. The objective for adaptive modulation becomes :

$$\text{maximise} \sum_{s=1}^{S} M_s, \tag{5.7}$$

$$\text{s.t} \quad \bar{b_e}(M_s) < b_t, \tag{5.8}$$

$$M_s \in \left\{ \begin{array}{cccc} 0 & 4 & 16 & 64 \end{array} \right\}, \tag{5.9}$$

where the averaged uncoded BER $\bar{b_e}$ of K resource blocks in (5.8) is given using the tight bound [108]

$$b_e(M_s) = \frac{4(\sqrt{M_s}-1)\sum_{m=1}^{M} Q(\sqrt{\frac{\frac{S}{N}(s,m)}{M_s-1}})}{\sqrt{M_s} M \log M_s}, \tag{5.10}$$

where $Q(x)$ is the Q-function, $Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} e^{\frac{-t^2}{2}} \mathrm{d}t$. (5.8) provides the average BER for the M resource blocks, indexed as $m = 1..M$. The factors $\frac{4(\sqrt{M_s}-1)}{\sqrt{M_s} M \log M_s}$ appears as a result of the union bound [108].

The maximising values of M_s for (5.7)-(5.9) are computed by performing a search over M_s and plugging the evaluated SNRs into (5.10) and further verifying that the BER target b_t is met in (5.8). Note that because the uncoded BER is averaged across all the frequency resource blocks,

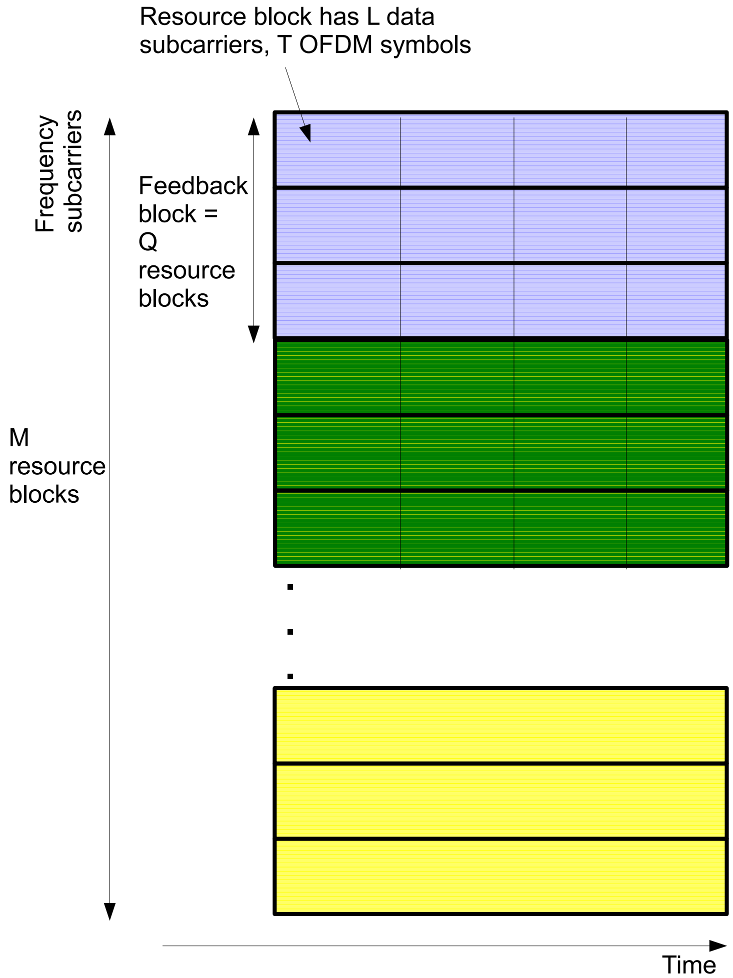


Figure 5.4: Description of feedback block. The example shows 3 resource blocks to make a feedback block.

it implies that the coded symbols are mapped across the frequency resource blocks. However, we do not account for the gain provided by this frequency domain mapping to the Viterbi decoding algorithm, as pointed out in [111] because of the computational cost at user equipment.

5.4.2 Fine feedback

The fine feedback scheme increases the frequency granularity in applying adaptive modulation so that the resource blocks are used more efficiently. A user equipment informs the relay about the adaptive modulation for pre-grouped bundles of Q resource blocks, which is termed a feedback block as shown in Figure 5.4. Increasing the number of feedback blocks clearly increases the amount of feedback data. All the feedback information cannot be sent through PUCCH and therefore needs to be sent via the uplink data traffic channel. In effect, this would compromise the achievable uplink data rate.

The search for the best modulation order is performed by varying the MIMO rank per feedback block. For a given MIMO rank, the SNR value per spatial stream is obtained with the MMSE equaliser for all the resource blocks within a feedback block. The option of transmit diversity is incorporated through antenna selection. Now the task is to calculate the highest modulation order that is supportable on a feedback block and for a given MIMO rank. The method used for coarse scheme, in (5.7)- (5.10), can be applied with an average BER target in a feedback block.

However, we take a slightly different approach so that the resource blocks can be used with more scalability. This means that a user can be assigned any number of resource blocks ranging from l to M without affecting the target BER. One scenario is that it can improve the utilisation of resource blocks for small payloads.

The following scheme is used per feedback block : the modulation m_{si} per spatial stream s is obtained for each resource block i, with a SNR mapping as in Table 1 which has been drawn for a target uncoded BER of 10^{-3}. The user equipment signals m_s as the adaptive modulation for the feedback block simply as

$$m_s = \min_{i=\{1..Q\}} m_{si} \tag{5.11}$$

i,e it chooses the minimum value of m_{si} in Q resource blocks. [1]

In doing this, the best MIMO rank for the feedback block is also signaled as the one which maximises $\sum_{s=1}^{s=S} m_s$, i.e we maximise the minimum modulation order on a feedback block. A fixed error correction code is used on all the resource blocks to obtain the PER from the BER.

Consider the design for PER less than 10^{-6} which can be used for TCP applications [106]. The uncoded BER of 10^{-3} that is used to draw Table 5.1 along with a $\frac{5}{6}$ convolutional code can be shown to achieve this PER in AWGN channel for 2500 byte payload and using a Viterbi decoding

[1]Note that for small values of $Q < 4$, (5.11) provides a good trade-off between the granularity and performance of feedback. In addition, indoor relay scenarios are also short range with low delay spread and high correlation between the resource blocks

algorithm. These values are obtained with simulations and further verified by the results in [110]. It is remarkable that a high channel code rate can achieve very low PERs in the AWGN channel if the uncoded BER is low. We exploit this gain in the fine feedback scheme and apply a channel code $\frac{5}{6}$ across the time domain on each resource block. In comparison, for the previous coarse feedback method, the code rate is applied across the frequency domain after setting the average uncoded BER target to 10^{-3}.

5.4.3 Extended feedback

The extended feedback scheme evaluates the SNR in a method similar to the fine feedback scheme, but uses more channel coding and matrix precoding options. The ultimate aim is to utilise more granular SNR values through channel coding. The utilisation of the coding rates is in turn increased using spatial precoding. To this end, adaptive channel coding defines SNR regions in between the values corresponding to adaptive modulation. Closed-loop MIMO precoding improves the achievability of the SNR regions. Adaptive modulation and coding is performed after choosing the best precoder out of a pre-defined set of precoding matrices for each MIMO rank. The choice of best precoding and rank combination is informed to the relay. Thus feedback information from the user equipment signals the choice of best precoding, MIMO rank, and as well as adaptive modulation and coding. The mapping table is shown in Table 5.1, where additional code rate options of $\frac{2}{3}$ have been inserted for the target PER of 10^{-6}, based on SNR values which are obtained from simulations and consistent with [110]. For codebook based precoding, 3GPP LTE standard defines a unitary codebook set based on the number of transmit antennas. For the 2×2 MIMO, the codebook set

Table 5.1: Modulation mapping. Ext-Extended

SNR	Modulation	Code rate	Schemes
≤ 7.5 dB	No transmission	No transmission	Ext, Fine
> 7.5 dB	QPSK	0.667	Ext
> 9.5 dB	QPSK	0.833	Ext,Fine
> 13.5 dB	16-QAM	0.667	Ext
> 16 dB	16-QAM	0.833	Ext, Fine
> 19.5 dB	64-QAM	0.667	Ext
> 22.5 dB	64-QAM	0.833	Ext, Fine

$\left\{ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}, \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ j & -j \end{bmatrix} \right\}$ is used. A concise summary of the number of feedback bits

Table 5.2: Number of feedback bits for wideband spatial multiplexing: 2 × 2 MIMO, 20 MHz

Feedback Type	Link adaptation	MIMO	Modulation+Rank	Coding	Precoding	Total bits
Coarse	Wideband	open-loop	4	0	0	4
Fine	Granular	open-loop	64	0	0	64
Extended	Granular	closed-loop	64	32	32	128

along with different functionalities is shown in Table 5.2 for 2×2 MIMO, with 48 resource blocks in 20 MHz bandwidth, and 3 resource blocks making a feedback block ($Q = 3$).

5.5 Downlink coverage results

5.5.1 The measurement environment

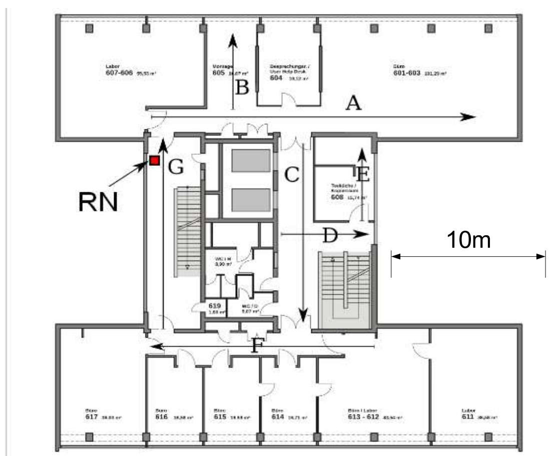


Figure 5.5: Indoor floor plan of the measurement site

The measurements were conducted at the 5^{th}, 6^{th} and 7^{th} floors of the Heinrich Hertz Institute office building. The layout of the office in 6^{th} floor (at height approximately 20 m) is shown in Figure 5.5. The office consists of corridors and the measurements were taken by moving a user equipment on a small trolley along the pre-defined paths shown in the corridors. The office is equipped with standard office furniture and has doors at sides of the floor while one of the doors consists of a glass front. On each floor, the office layout as shown in the same plot, Figure 5.5 consisted of paths A to

G, approximately of lengths 20 m, 5 m, 10 m, 5 m, 5 m, 10 m and 10 m respectively. Paths A, C, F and G are corridors. Paths B, D and E are small sections opening from the corridors. Corridor G is closest to the macro base station which is at a distance of approximately 500 m. Few sections of the corridor G receive a line of sight signal from the macro base station through a window. The windows are not heavily coated, which allowed for high received signal power in its vicinity. The start of the measurement was from a door at the start of track A, going through on full trip away and the end was at the meeting point of tracks A and G.

5.6 Measurement Setup

A three terminal relay network, consisting of an outdoor base station (BS), an indoor relay node (RN) and an indoor user equipment (UE) was set up for measurements. All the links are 2×2 OFDM-MIMO with feedback from each receiver to its transmitter. A bandwidth of 20 MHz is used at 2.6 GHz carrier frequency. No other BS or UEs were active during the experiments. The aim of the experiment was to characterise the gain of relaying for outdoor to indoor downlink coverage in a typical urban cellular indoor environment. The measurements used the parameters shown in Table 5.3, with the macro base station using a per antenna power constraint of 40 dBm and a uniform transmit power spectral profile. The base station antennas were sectorised with 11 dBi directivity gain towards the measurement site. Cross-polarised antennas, as in Section 2.2.2 were used at the base station so as to employ polarisation multiplexing to the receiving terminal (which can be the relay or the user equipment) as was shown in [121]. Similarly cross-polarised antennas were employed at the transmitter of relay access unit for the link to the UE. For channel quality reporting, the fine feedback scheme explained in Section 5.4.2 was implemented with a target uncoded BER of 10^{-3}. This feedback scheme is implemented independently for the two hops; base

BS, RN transmit powers = 43 dBm, 23 dBm	BS downtilt = 10°
RN isolation = 20 dB	RN feeder-access units duplexing = TDMA
Carrier frequency = 2.6 GHz	Transmission bandwidth = 20 MHz
Receiver detector = MMSE, MRC	AGC = per RX antenna
BS, RN transmit antennas = 2 xpol	UE, RN receive antennas = 2 xpol

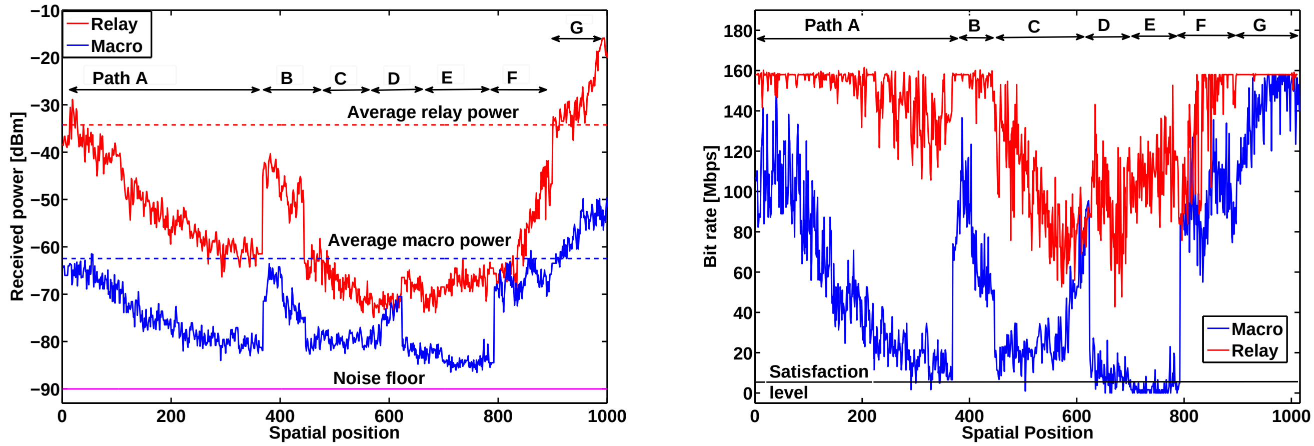
Table 5.3: Downlink transmission parameters of indoor measurements

station to the relay and relay to the user equipment. The achievable bit rate is characterised using (5.6) with values of $N_d = 144, \beta = 0.95, T_s = 0.5$ ms.

The terminals were set up as follows. The relay feeder unit, which connects bidirectionally to the base station was positioned in corridor G on the 6th floor, and close to the window, and facing the base station. The relay access unit, which connects bidirectionally to the user equipment, was positioned indoors in corridor G and used two isotropic antennas at a height of 3 m. An ethernet cable connects the relay feeder and access units. The office is equipped with standard office furniture and has doors at sides of the floor while one of the doors consists of a glass front. Note that in case of heavy window coating, the relay feeder unit's antennas will be placed outside the window but the access unit's antennas can still be placed indoors to avoid the penetration loss. The relay receiver, the relay transmitter and the user equipment receiver terminals used two cross-polarised antennas each to make use of polarisation multiplexing, wherein the two MIMO data streams are fed into a separate polarisation mode. Polarisation is also found to be useful for outdoor range extension, as we have shown in some more results in [117].

Four downlink measurement runs were performed. The first measurement was a direct outdoor to indoor measurement in the 6^{th} floor with the relay node manually switched off. Later three more measurements were collected in the 5^{th}, 6^{th} and 7^{th} floors with the DF relay transmitting from the 6^{th} floor. The relay coverage can thus be compared to the macro coverage on the 6th floor.

During the measurements, a multimedia streaming application was transmitted from the base station and routed via the DF relay to the UE. For routing the video content, network layer relaying, also known as layer 3 DF relaying is implemented. A layer 3 relay works by using the decoding and forward protocol on an Internet protocol (IP) packet basis. At the base station the multimedia content is encapsulated into IP packets and numbered. The IP packets are then segmented into transport blocks which are further mapped to channel resource blocks and transmitted. The UE is assigned an IP address at the base station. The DF relay uses the same IP address for forwarding the packets and can discard the packets which are not meant for the UE. The relay transmits the successfully received IP packets after a renumbering mechanism. This mechanism avoids redundant transport block retransmission to the UE in the case of packet errors. It is noted that eventhough



(a) Received signal power

(b) Bit rate

Figure 5.6: Measurements comparing coverage with and without a relay using 2 $\times$ 2 MIMO in OFDM for 20 MHz bandwidth

very low PER is targeted without HARQ, to still correct any packet errors occurring at the target PER, HARQ work independently on the two hops.

5.6.1 Outdoor to Indoor Macro Coverage

Real-time measurements from the macro BS to the indoor site are shown in Figure 5.6a and 5.6b. Figure 5.6a shows the received power and Figure 5.6b shows the measured bit rate, in the 6th floor and using the fine feedback scheme in Section 5.4.2. The X-axis shows the spatial position in terms of the sample index in the measurement path. In the corridor G high signal power and bit rate were received because of the line of sight to the base station. Similar bit rates were observed in parts of corridor F and A which are close to corridor G. However, approximately 2.5 m away (sample 100) on corridor A, a sharp decline is observed. At the end of corridor A, 10 m away from the start, bit rate has reduced to 5 Mbps.

The bit rates in path C were more or less the same as compared to the end of corridor A. These two corridors are connected via a glass door to corridor A which validates such an observation. Path D showed some more degradation in bit rate to less than 4 Mbps. However, path E is behind another wall and showed deterioration to less than 2 Mbps because of the wall penetration loss of about 10 dB.

Typically 2 Mbps is taken as a satisfactory level for bit rate [11]. Some locations in A, C, D and E showed bit rate less than this satisfactory level, and are therefore said to be coverage holes. The ratio of coverage holes amounted to approximately 8% of all the measured paths, i.e, LTE direct link suffers from 8% indoor coverage hole probability. However, a user experience can be even worse. This occurs because the coverage holes are not uniformly distributed. In the count of measured paths, the 2 paths, D and E out of the total 8 paths showed a higher probability of containing a coverage hole. Thus approximately 25% of the total measured area cannot be guaranteed a satisfactory bit rate level with high probability. The problem is that a user who is unaware of the coverage pattern in locations D and E may conclude that there is no connectivity. Clearly, a user cannot be expected to know the spots in the problematic locations which are ideal for installing a laptop. To worsen the situation, 20% of the paths D and E showed zero bit rate. These observations are remarkable considering that only 10 m away on the other side on the same floor, bit rate of 100 Mbps was measured. In this case, the signal to noise ratio (SNR) averaged over all the spatial locations was 27.5 dB. It is however possible that places with very low received power could potentially be served with a more robust code rate. This possibility is demonstrated using the extended feedback scheme during the off-line analysis which is shownlater in this Chapter.

5.6.2 Coverage with a relay

The received signal power and bit rate measurements with a relay are shown in Figures 5.6a and 5.6b respectively for the same floor (6th floor). In Figure 5.6b, the bit rate measurements are shown for the indoor downlink using 20 MHz, i.e, between the indoor relay and the user without accounting for the feeder link bandwidth as modeled in (5.5). Relaying improves the received signal power, from an average of -60 dBm from the outdoor base station to -30 dBm from the indoor relay. A correspondingly higher bit rate exceeding 50 Mbps is also received from the relay. Importantly the bit rate improvement was achieved using a transmit power of 23 dBm which is significantly lower than the 43 dBm used by the macro. In the case of a relay, Figure 5.6a shows a variation of signal power from -20 dBm to -70 dBm. This power fluctuation can be exploited for feedback reduction as we note later. However, the fluctuation of bit rate is more for the macro link. Thus the

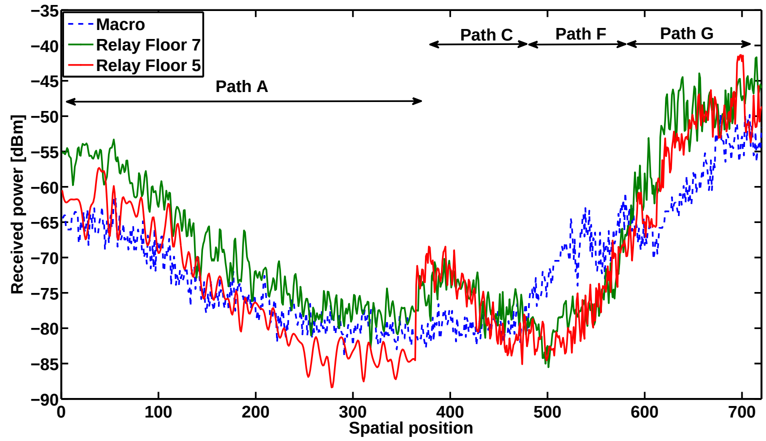


Figure 5.7: Relay coverage in different floors

20 MHz macro downlink can be seen to be susceptible to sharp coverage variations. This happens because resource blocks which realise a SNR less than 9.5 dB were not utilised.

The relay worked on a more liberal SNR link budget of 15 to 70 dB SNR which overcomes the need for frequency diversity. In relay link, the signal to noise ratio (SNR) averaged over all the spatial locations was 57 dB. During the measurements, the relay feeder unit was placed in corridor G and received a high signal power of -50 dBm and a peak bit rate from the macro BS. But in reality the feeder unit can be placed anywhere in paths A, B, G and F. Because the bit rate from the outdoor base station showed a gradual decline along a corridor, even if the feeder unit were to be placed in any other indoor corridor, one can expect robust results from relaying. The received signal power on the floors 5 and 7 from the relay on floor 6 is shown in Figure 5.7. The signal strength from the relay was comparable in floors 5 and 7 but again varied spatially in the corridors of either floor. In certain sections of the floors 5 and 7, the macro signal power was considerably higher than the relay signal power. Thus selection handover between the relay and macro BS can provide gains around 5 dB in the overlapping coverage areas of macro and relay.

5.7 Impact of Feedback on Coverage

In Figure 5.8, numerical result of the link data rate is shown via offline simulation for different feedback schemes that were discussed in Section 5.4. The distribution of link data rate from the relay to different indoor spatial locations (1000 samples) is shown. The median data rate is measured at the 50 percentile point of the cumulative distribution function (cdf) shown in Figure 5.8.

Coverage improvements are represented by improvements in coverage data rate and minimum data rates, which are defined as follows. Coverage date is measured at the 5 percentile point of the cumulative distribution function of the data rates collected within an indoor cell. The minimum data rate is the tail of the cdf curve, i.e the data rate corresponds to the zeroth percentile. All data rates are calculated by applying (5.6).

From the relay-to-user link data rates, the end-to-end efficiency of relaying from (5.4) can be deduced by taking into account the time sharing between the relay's feeder and access links, and the feeder link efficiency. In practice depending on an operator's choice of feeder bandwidth and the feedback scheme, the effectiveness of relaying may vary, as is evident from (5.1) - (5.5). To observe the effect of feedback, the data rate distribution with and without a relay are compared for different feedback qualities in Figure 5.8 using a target PER of 10^{-6}. For these simulations, the stored set of channel matrices from the measurements on floor 6 were used. For consistency, we ensure that the simulations with the fine feedback scheme closely fit the real-time bit rate measurements shown in Section 5.6.1. A slight deviation can occur because the real-time measurements underwent impairments such as phase noise error which is not modeled to generate Figure 5.8. Results show that in the median the coarse feedback scheme behaves almost as well as the fine feedback scheme for the macrolink coverage (without a relay). The averaging effect over the frequency selective channel in the coarse scheme compensates for the lack of granular feedback. A significant gain is obtained from the fine feedback scheme for minimum data rate. Even though the minimum rate is zero for both the schemes, the probability is just 7 % in the fine scheme as compared to 40 % in the coarse scheme.

The extended scheme improves both the minimum and median data rates from the outdoor base station. This shows the effectiveness of closed loop precoding for the macro link, especially when more channel coding options are available. But the downlink gain may come at the cost of uplink loss. The implication on the uplink can be better understood in the light of voice over IP simulation

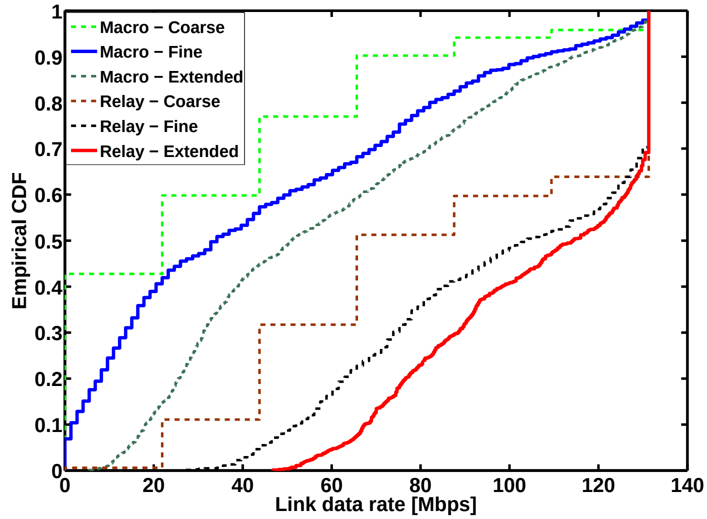


Figure 5.8: Performance comparison of various feedback schemes in distribution of data rate using 20 MHz bandwidth. The better performing curves are to the right.

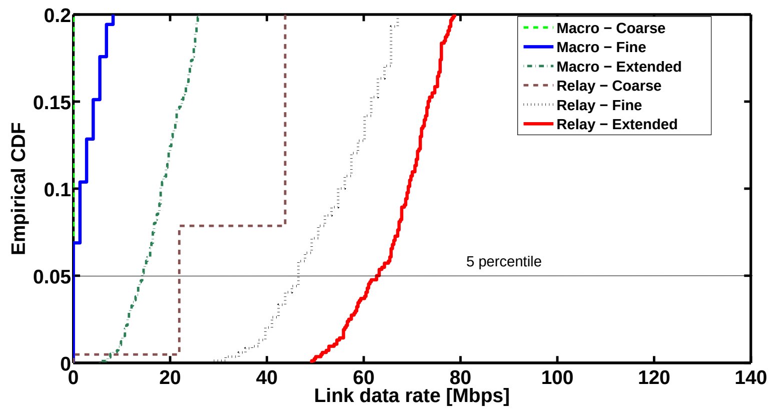


Figure 5.9: Lower percentile data rate region zoomed from Figure 5.8. The better performing curves are to the right.

results shown in [107]. The uplink satisfaction chances are seen to drop sharply with path loss for a 42 byte payload and for 200 users in a cell. Thus it may be challenging to support several voice over internet protocol (VoIP) users who are indoors at the same time.

The 128 bits of extended feedback may be encoded to 32 bytes with $\frac{1}{2}$ rate channel coding. If the feedback were to be updated once in 40 ms, then 32 byte of feedback comparable in size to the 42 byte VoIP payload. A macrocell may prioritise VoIP data in the uplink, and hence compromise the feedback quality. In the case of a relay, the 30 dB average SNR gain can be used for more feedback bits. Suppose there are only few resource blocks available for feedback. An indoor user can communicate more feedback bits to the relay by using higher order modulation. For example, instead of QPSK, 16 QAM modulation can be used on the feedback bits.

Based on this viewpoint, relaying has better chances of employing the extended scheme (red curve in Figure 5.8). In Figure 5.9 we observe that closed-loop precoding on the indoor channel increases the coverage data rate (5 percentile of the cdf) from 47 Mbps to 62 Mbps, which is a 32% gain. The minimum achievable data rate increases from 30 Mbps to 45 Mbps which is a 50% gain. The main benefit from this access link improvement is that a higher feeder to access bandwidth ratio Δ can be utilised to attain the same spectral efficiency of (5.4). The higher feeder bandwidth in turn will provide a better user efficiency U_r as can be seen in (5.5).

With an indoor relay, the coarse level feedback is able to achieve 60 Mbps in the median i.e, a spectral efficiency of 3 b/s/Hz. This result is significant for deployment, because the peak data rate is achievable in 35% of the locations even with the coarse feedback scheme. A two-pronged approach will therefore improvise: a) for median data rate, employ coarse feedback based on user selection and b) for minimum data, increase the feedback to an extended level.

5.8 Uplink limited feedback for multiple users

Previous section considered a single user. We now elaborate on the multi-user scenario in which the amount of feedback is constrained by the uplink capacity. In this case communicating the feedback data of all the users may not be feasible. Thus we do the following: a) propose a selective feedback scheme so that the amount of feedback is less than the maximum limit and b) apply this feedback scheme for relay downlink coverage results.

5.8.1 User selective feedback scheme

Assume that each user could choose among v feedback levels and the sum of all feedback overheads can not exceed J bits. We use the following variables : the choice of feedback level of user k is denoted using e_{kv} and n_v is the number of bits in a feedback level (uplink overhead). Clearly, depending on the feedback level, a corresponding MCS on a resource block is possible. Moreover, the chances of allocating a particular resource block varies with the MCS. Accordingly, let us denote the MCS of v^{th} feedback level using u^v_{km} and the corresponding allocation with variables x^v_{km}. A summary of variables along with a numerical example is given in Table 5.4. Now consider the following proportional fair objective

Variable	Description	Our example
v	Feedback level	$v = 0, 1, 2$ denotes coarse, fine and extended
u^v_{km}	MCS of feedback level v	Constants, Refer Table 5.2
n_v	Number of feedback bits on uplink data channel	0, 64, 128 bits for $v = 0, 1, 2$ respectively
e_{kv}	Feedback level allocation variables v to user k	Binary values 0 or 1
x^v_{km}	Resource block allocation variables	Binary values 0 or 1
J	Total feedback bits (over all the users).	75% of fine feedback scheme.

Table 5.4: Table of variables and constants

$$\max LT \sum_{k=1}^{k=K} \log(\sum_{m=1}^{m=M} \sum_{v} e_{kv} x^v_{km} u^v_{km}) \tag{5.12}$$

$$\sum_{v} \sum_{k} x^v_{km} = 1 \quad \forall m \tag{5.13}$$

$$0 \le x^v_{km} \le 1 \tag{5.14}$$

$$\sum_{v} \sum_{k} n_v q_{kv} \le J \tag{5.15}$$

$$e_{kv} \in \{0, 1\} \quad \forall k, \forall v. \tag{5.16}$$

We thus need to obtain x_{km} and e_{kv} $\quad \forall v, k, m$. The two problems are difficult to solve , even if the feedback variables e_{kv} are to be continuously relaxed. To focus on the problem of feedback allocation, we may assume that all the x^v_{km} equals to $\frac{1}{K}$. The problem is then the following

$$\max LT \sum_{k=1}^{k=K} \log(\sum_{v} e_{kv} \sum_{m=1}^{m=M} u_{km}^{v}) \tag{5.17}$$

$$\sum_{v} \sum_{k} n_v e_{kv} \leq J \tag{5.18}$$

$$e_{kv} \in \{0, 1\} \quad \forall k, \forall v. \tag{5.19}$$

The above objective now demands that higher feedback level be allocated to the user who benefits the most. Based on this logic, we provide binary integer solutions of e_{kv}.

The main difficulty is however to relate the benefit of a feedback level to its uplink cost n_v. Thus, for example, we may have to choose between allocating 128 bit feedback to one user or 64 bits to two users. To find a trade-off between the benefit and the cost, we propose the following scheme.

5.8.2 Proportional fairness with selective feedback

To begin with, we demarcate the set of users into two sub-groups. One group of users, for which coarse feedback achieves the peak downlink data rate. The second group of users for which coarse feedback does not achieve the highest rate. The following two-step method is further applied for the second group of users.

1. Initial allocation : Select a user k^* and feedback level v^*+1 based on $\arg\max_v \max_k \frac{\sum_m u_{km}^{v+1} - \sum_m u_{km}^{v}}{\sum_m u_{km}^{v} \times (n_{v+1} - n_v)}$. This means $e_{k^*(v^*+1)} = 1$ It is noted that this feedback level is applied on all the M resource blocks. The best user k^* is removed from the pool of users for feedback level $v^* + 1$ and the feedback capacity is recalculated. This process is repeated until the the total J feedback bits are exhausted.

2. Iterative swapping : Check whether all the users (in the so-called second group of users) have been allocated at least one feedback level greater than coarse ($v = 0$) in the previous step. If not, swap a pre-chosen feedback level v' from user k' to a set of users K^l. These users are given a pre-chosen feedback level l each, s.t $\sum_{K^l} \log(\sum_m u_{km}^{l}) - \sum_{K^s} \log(\sum_m u_{km}^{v=0}) \geq$

$\log(\sum_m u_{k'm}^{v'}) - \log(\sum_m u_{k'm}^{v=0})$ and $|K^l|n_l = n_{v'}$. This means user k' goes to coarse level.

After allocating feedback levels to users, we need to perform resource block allocation. It is noted that for users with coarse feedback level i.e, $v = 0$, equal time sharing is optimal for the proportional fair objective. This fact applies irrespective of the allocation that is done for users with higher feedback levels. Thus the users can be first split into two separate user groups. One group consisting of users with coarse feedback (which is frequency flat) and another consist of frequency selective feedback. Based on the above observation, we implement the following scheme.

1. Create two user groups K_{coarse} and K_{freq}; K_{coarse} the sets of users allocated coarse feedback level and K_{freq} the set of users with any other feedback level other than coarse. Let k_{coarse} and k_{freq} denote the number of users in K_{coarse} and K_{freq} respectively.

2. Apply equal time sharing for the users K_{coarse}.

3. Perform proportional fair resource block allocation for the user set K_{freq}.

4. Now obtain optimal time sharing between the two groups K_{coarse} and K_{freq}. The optimal ratio of time slots for group K_{coarse} is $\frac{k_{coarse}}{K}$. This ratio of time slots is to be further shared evenly between k_{coarse} users. Thus each coarse user gets a time share of $\frac{1}{K}$.

5. The effective time slot allocation for users in K_{freq} is $1 - \frac{k_{coarse}}{K}$. The data rates of users in K_{freq} are now scaled down by this factor.

5.8.3 Improvement with selective feedback

This section demonstrates through simuation results that selective feedback scheme improves the proportional fairness level that is achieved. One could infer from the proportional fairness metric in (4.4) that proportional fairness prevents deprivation of data rates to users with lower MCS levels. These low data rate users also correspond to the 5 percentile region of the cumulative distribution functions. Thus proportional fairness metric is a good indicator for coverage data rate and minimum data rates. Results for the proportional fair rate sharing with selective feedback is

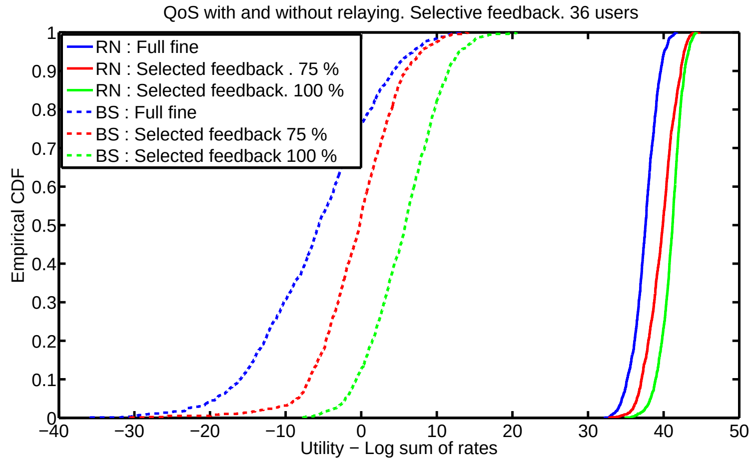


Figure 5.10: Performance of user selective feedback for proportional fairness, shown with and without relaying. The results demonstrate the effectiveness of feedback allocation based on user location.

shown in Figure 5.10. We assume a simple scenario of 36 users, and the three feedback levels as in Table 5.1. In the case of relaying, the result is shown for the relay to user access link without taking into account the feeder link. This is a fair comparison because we wish to show that a relay can make better employ different feedback levels on its access link. From the plot, we see that the relay access link hardly suffers any loss (less than 5%) when the amount of sum feedback (i.e J) is reduced to $0.75 \times 36 \times 64 = 1728$ bits rather than $36 \times 64 = 2304$ bits. The coarse feedback level has zero overhead because this feedback will be communicated on the uplink control channel instead of the uplink data channel. Selective feedback with total of 1728 bits performs even better than employing fixed 2304 bits of fine feedback (64 bits per user) across the 36 users.

In the case of direct downlink (dotted lines), the impact of feedback capacity is more pronounced. Here a higher feedback capacity of 2304 bits along with selective feedback on the uplink improves the performance tremendously. However, this improvement is still not significant enough as compared to the gain obtained with a relay.

Thus we see that a relay has in fact a double benefit, a) improves the downlink quality of service objective and b) allows us to reduce the feedback overhead that is needed to achieve a high quality of service.

5.9 Conclusion

This chapter presented a relay deployment concept using decode and forward relaying. Field trial results of relaying were conducted with a real-time LTE test-bed in an indoor office environment of dimensions 20 $m \times 10m$. A very low PER and only a limited feedback of four MCS levels were used. Furthermore, coverage with a relay was compared to the direct macrocellular link without a relay in indoor locations. Macrocellular measurements from outdoor to indoor showed the existence of coverage holes. The variation of data rate can be sharp, with a change from 100 Mbps to 5 Mbps within 10 meters. Spots with zero data rate were also observed. Measurements from an indoor relay shows full coverage and substantial data rate improvement.

To analyse the effect of feedback, a classification of feedback levels were made as coarse, fine and extended. In the case of a relay, coarse feedback, which involves no precoding and employs only wideband MCS is seen to be sufficient for median and peak data rates. The minimum data rate may however be quite low (less than 1 Mbps). Fine feedback, which performs frequency dependent link adaptation increases the minimum data rate to 30 Mbps. This represents a substantial improvement as compared to the coarse feedback. A further improvement in minimum data rate to 45 Mbps (a 50% gain) is obtainable using the extended scheme, which employs spatial precoding at the relay.

There may be practical constraints in a system regarding the amount of feedback that can be sent on the uplink. This limitation is modeled using an upper limit to feedback capacity. A feedback switching mechanism, called user selective feedback is proposed to deal with such scenarios. The basic idea of the scheme is to allocate one of the feedback levels, coarse , fine or extended to a user based on the utility to that user.

Finally, we extend relaying to multiple users, and illustrate results of proportional fairness on the relay link. Relaying suffers only a loss of less than 5% when the sum feedback from all the users is reduced by 25%. The direct macrodownlink however undergoes a substantial loss. These results correspond to non-guaranteed bit rates in a single relay cell. In the next chapter, we consider relaying for guaranteed bit rates to multiple users, with multiple relays in a macro-cell.

Chapter 6

Multi-user multi-relay network

6.1 Introduction

Relaying can be used to improve the coverage of many users at the same time. A particular scenario for relaying which we foresee in future networks is delivering streaming applications to many indoor users.

Thus we look into the question if relaying is beneficial for these applications when the downlink is bandwidth constrained. A natural problem which might arise is that a half duplex relay sacrifices half the degrees of freedom. As pointed out in early works, this straightaway penalises the spectral efficiency of the relay link, thereby counteracting against the benefit of higher signal power.

Here we consider cellular relaying by incorporating two possibilities. Multiple users per relay and multiple relays per macro cell. The multi-user transmission is enabled by using OFDMA in the relay access link. Multiple relay deployment is made through spatial reuse (where relays use the same frequency by exploiting spatial separation). We consider the scheduling problem for guaranteed user data rate under this relaying architecture. We call relaying framework guaranteed bit rate by relay scheduling (GBRS).

Our results show that eventhough there is a half duplex constraint on relays, it can be effectively countered from the gain of scheduling and spatial reuse. In particular the solutions we propose employ hierarchical user scheduling and admission control that increases coverage of GBR applications.

Numerical results are shown in an indoor scenario using two sets of channels from a:) a direct link from a macro base station 500 m distance and b) indoor measurement with a decode-and-forward relay. Significant improvement in the number of GBR users is shown by employing the proposed concepts.

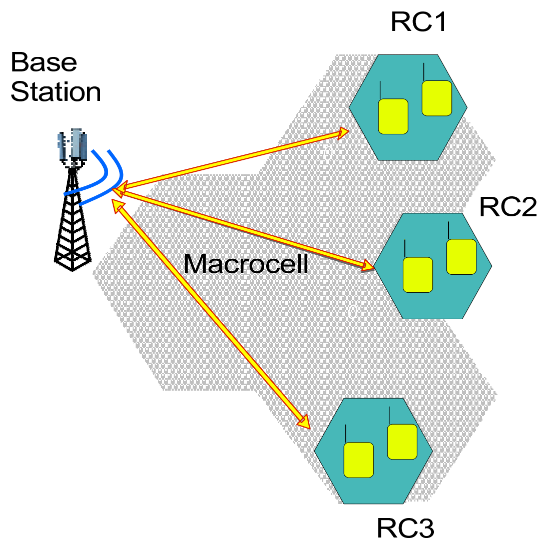


Figure 6.1: A set-up of multiple relays in a macrocell

6.2 System model and assumptions

We consider a relaying scenario as shown in Fig 6.1. A macro base station supports many users in the downlink OFDM air interface. We employ frequency division duplexing (FDD) of the uplink and the downlink. The downlink air interface consists of M resource blocks for a time slot duration, with each resource block consisting of a number of data symbols. A data subcarrier carries a data symbol over a duration of one OFDM symbol [2]. The base station uses a sum downlink transmit power of P_o W and employs uniform transmit power profile across the subcarriers. The users are

[2]We simply say 'bits per subcarrier' instead of 'bits per data symbol on that subcarrier' for brevity and yet emphasise frequency dependent MCS

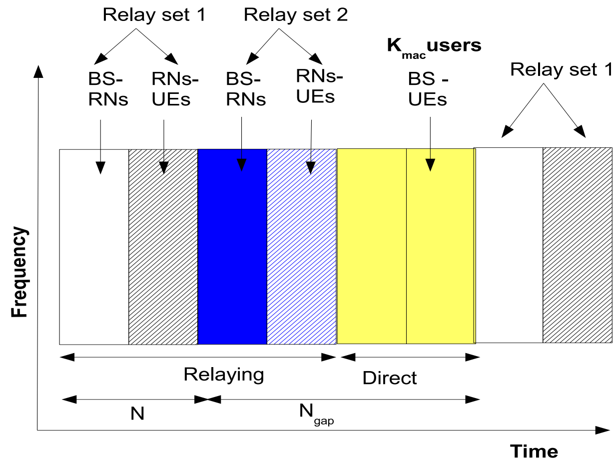


Figure 6.2: Frame diagram showing dedicated time slots for relaying

assumed to be situated indoors. Each indoor site receives low signal power from the base station and is thus equipped with a infrastructure helper node called 'relay'. The term feeder link refers to the connection between the base station and relay, while access link refers to the connection between the relay and user. The feeder link is realised to a feeder unit of the relay, while the access link is supported on a separate access unit. The two units are interconnected using an ethernet cable. There are thus two pairs of FDD links, base station to relay and relay to users. There are R such relays, with each relay connecting to a different set of users. Each relay employs OFDMA in the access link, so that multiple users can be served in a single time slot. The base station uses OFDMA to send feeder link data to multiple relays. Thus, we are in a situation of hierarchical OFDMA. The feeder and access links use the same carrier frequency and bandwidth.

Discussion on System level integration

The relays are deployed in macro-cells to improve coverage and hence some resource separation would be needed for relaying at system level. As shown in Figure 6.2 this separation can be done by setting aside time slots for relaying. Eventhough it might appear that relaying thus consumes 'extra' radio time slots and thus essentially degrades macro direct downlink, it is not the case. The reason is because the relay users are macro users who have been further handed over to the relays for better coverage. Importantly, the macro-base station still acts as the information source for the relay link. Thus in the absence of a relay, these relay users would indeed be directly connecting to that macro-base station and still consume those so-called relay time slots. The following cases will need to be considered at a system level.

1. Macro only users-Users in the cell who receive a higher signal to noise ratio (SNR) from the macro-base station (as compared to relay) are denoted as 'macro only users'. They are directly served by the macro-base station. These users, shown as K_{mac} in the figure are served in the radio time slots meant for macro only users.

2. Time division of relay sets- Many relays may operate with spatial reuse (full frequency reuse) by exploiting their smaller cell sizes. However, there may be instances when there are adjacent relays on the same frequency, e.g, an operator may have installed relays on adjacent floors. In such situations, one relay may cause interference to the users of another relay. To handle such cases, it is taken that non-interfering relays form a set and are scheduled together in a time slot. Thus a set of relays which interfere is scheduled in a different time slot as in Figure 6.2, where relay set 1 and relay set 2 are different sets of relays.

3. Macro to relay interference - Eventhough it is assumed that macro base station does not serve the macro only users during the relay slots, it still needs to be active in order to deliver the feeder data to relays. Thus macro to relay interference may still take place. To handle such instances, macro power control is an effective functionality, which we show and discuss in our results later in this chapter.

Based on the above, it is noted that after taking into account the macro only users and relay set time division, there can be a time gap between two servings of a relay set, denoted as N_{gap} in Figure 6.2. Thus at the system level a scaling factor of $\frac{N}{N+N_{gap}}$ needs to be accounted for bit rate calculation. However we observe that this scaling factor would be applied even without relays (if the macro serves those users with a time gap). The main problem thus becomes whether the relays can utilise their time slots efficiently as compared to the macro-direct downlink. In the remainder of the thesis, we focus on this core problem of efficient usage of time slots by a set of relays. Through this we wish to show that our relay schemes are effective in the usage of the so-called 'relay time slots' and can support more guaranteed bit rates as compared to the macro-direct downlink.

Relay nodes abide by the half duplex constraint, and therefore do not transmit and receive at the same time slot. Each relay node can use its allowed maximum transmit power of P_r W. Multiple relay cells (RC) can be simultaneously active within a macro cell, such as RC1, RC2 and RC3 as shown in Figure 6.1. Even though, such an assumption may at first appear to be optimistic, the following reasons justify the case : a) in typical indoor environments, the building penetration loss is significant as shown in the measurements in [118] and adds to the path loss. Thus the interference between any two relay cells can be mutually well degraded w.r.t the signal power b) a base station is already aware of the positions of the relays (through global positioning system) and may schedule only well separated relays in the same time slot. We may thus safely assume that the relay cells are out of the coverage zone of each other. The feeder unit is suitably placed to receive high signal power from the base station, for example placed outside the window. We denote the average spectral efficiency that is supportable of the feeder link as γ_r bits per subcarrier. We use Δ_r to denote the number of resource blocks allocated to the feeder link of the relay r. There are K users indexed $\{1, 2, 3..K\}$ and uniquely mapped to R relay cells $\{1, 2, 3..R\}$. Thus a relay r supports k_r users. The set of these users for relay r is denoted as K_r. The transmission interval from base station to user is a fixed N time slots. Without loss of generality, it is considered that the k_r users are shadowed, and receive low signal power from the base station and are therefore in need of a relay. The indoor users who do not fall under this category are assumed to be directly catered by the base station in separate time frames as shown in Figure 6.2.

We illustrate the hierarchical OFDMA and full bandwidth reuse with an example in Fig. 6.3. Here, two indoor relay cells with many users are shown. $f1, f2, f3$ are feeder link bandwidths, e.g 1 MHz each, in the 2.6 GHz frequency band. A sum bandwidth of $f1 + f2 + f3$ is thus available to each relay for its access link, which we term full frequency reuse. A relay transmits only for $\epsilon_r N$ time slots out of N time slots. The base station, which uses $f1$ as the feeder link to relay 1 and $f2$ for relay 2, transmits for durations of $N - \epsilon_r N, \quad r = 1, 2$ to relay 1 and 2 respectively. We assume that the feeder link signal for relay 1 does not degrade the access link of relay 2 because of high

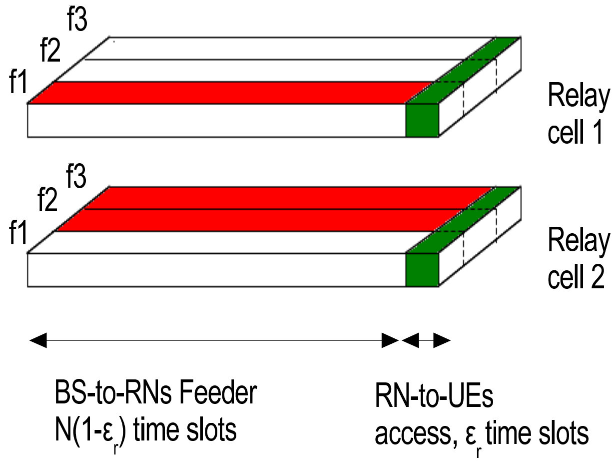


Figure 6.3: The framework for optimisation in relaying

attenuation loss. This interference situation may arise if $\epsilon_1 \neq \epsilon_2$. In addition, the base station may apply transmit power control on the feeder link to one relay so as to suppress any unwanted interference to the access links of other relays. In the optimisation problem, we wish to determine the following: optimal bandwidths for feeder links $f1, f2$, the optimal durations $\epsilon_r N$, the best set of users to be scheduled in each relay cell out of K_r , $\forall r$ and subcarrier allocation in the access link for the selected users.

6.2.1 Summary of variables

The following are the variables which are to be solved and the constants : Variables:

- x_{km} : Allocation variables showing the amount of allocation of m^{th} resource block to the k^{th} user in the relay to user access link. $0 \leq x_{km} \leq 1$.
- Δ_r : Allocation variables showing the amount of bandwidth in terms of resource blocks allocated to the r^{th} relay by the base station on the feeder link. $0 \leq \Delta_r \leq M$.
- ϵ_k : Duplex time sharing variables showing the time allocation to the k^{th} user's access link. This is normalised to N, the total number of time slots and are thus fractions.

Constants :

- M : The maximum number of resource blocks in the downlink for a macro-cell.
- u_{km} : The modulation and coding values for m^{th} resource block and the k^{th} user. These are obtained from the channel gains in the downlink. It is represented in loaded bits per subcarrier on an OFDM symbol (data symbol) and applied throughout that resource block.
- L, T : L is the number of data subcarriers per OFDM symbol in the downlink for a macro-cell. T is the number of OFDM symbols per time slot. The product $L \times T$ is for example 144 data symbols. The product $L \times T \times u_{km}$ will yield the bits loaded per resource block.
- N : The total number of time slots used end-to-end in a transmission frame between the base station and user.
- R : The number of relays in the macro-cell.
- γ_r : The spectral efficiency of the feeder link from base station to the r^{th} relay.
- K_r, k_r : K_r is the set of users indices affiliated to the r^{th} relay. The number of users in the set K_r is k_r.

6.3 Problem set up for inband relaying

We consider DF relaying, in which a relay fully decodes the data from the base station before forwarding to users. The relay applies uniform transmit power on all subcarriers, schedules the users on selected resource blocks, applies modulation and coding scheme (MCS) and re-transmits on the access link. Let u_{km} be the bits per subcarrier to user k on resource block m in the relay-to-user access link. Let x_{km} denote the fractional (between 0 and 1) allocation of resource block m allocated to user k and $N \times \Omega_k$, the number of bits to be loaded in N slots. The objective function for bit loading guarantees so as to maximise the number of users is given below :

$$\max \sum_k a_k \quad \text{s.t,} \tag{6.1}$$

$$a_k = 1, \quad \text{if} \quad NLT \sum_m u_{km} x_{km} \epsilon_k \geq N\Omega_k \tag{6.2}$$

$$a_k = 0, \quad \text{if} \quad NLT \sum_m u_{km} x_{km} \epsilon_k < N\Omega_k \quad \forall k \in \{1, 2, .., K\} . \tag{6.3}$$

In (6.3) ϵ_k denotes the fraction of time slots for which the relay access link to user k is active out of the total N time slots. We assume that the relay feeder and access links use the same set of frequencies, i.e, an inband relaying scenario as shown in Figure 6.3. Therefore the feeder bandwidth to a relay may overlap with part of the access bandwidth from that relay. This can result in an issue. When the relay access unit transmits on a resource block, it may cause severe interference on that resource block to its own feeder link (in spite of some isolation). With transceiver imperfections, this interference may leak to other resource blocks on the feeder link. Therefore, we enforce that when the access link of a relay r is activated to any affiliated user k of that relay, the feeder link data of all the users stops to that relay r. Because of this half duplex constraint, the variables ϵ_k should be the same for all users in a relay cell and becomes ϵ_r.

Under these conditions, the relationship between ϵ_r and Δ_r is deduced as

$$N\Delta_r \gamma_r (1 - \epsilon_r) = N\epsilon_r \sum_m \sum_{k \in K_r} u_{km} x_{km}, \tag{6.4}$$

(6.4) says that amount of feeder link data is equal to the amount of access link data in r^{th} relay. This gives the half duplex time sharing values,

$$\epsilon_r = \frac{\Delta_r \gamma_r}{\sum_m \sum_{k \in K_r} u_{km} x_{km} + \Delta_r \gamma_r} \quad \forall r \tag{6.5}$$

The objective in (6.3) can be relaxed using continuous variables a_k and then transformed with weights Ω_k^{-1} as we showed in Chapter 4. Upon plugging the value of ϵ_r from (6.5) into this relaxed objective, we get the following objective with resource constraints :

$$\max LT \sum_{k=1}^{k=K} \Omega_k^{-1} \frac{\sum_{m=1}^{M} u_{km} x_{km} \Delta_r \gamma_r}{\sum_{m=1}^{M} \sum_{k' \in K_r} u_{k'm} x_{k'm} + \Delta_r \gamma_r} \tag{6.6}$$

$$\text{s. t },\quad LT \frac{\sum_{m=1}^{M} u_{km} x_{km} + \Delta_r \gamma_r}{\sum_{m=1}^{M} [\sum_{k' \in Q_r} u_{k'm} x_{k'm} + \Delta_r \gamma_r]} \leq \Omega_k \quad \forall k, \tag{6.7}$$

$$\sum_{k \in K_r} x_{km} = 1 \quad \forall m, \forall r, \tag{6.8}$$

$$0 \leq x_{km} \leq 1 \quad \forall k, \forall m, \tag{6.9}$$

$$\sum_{r=1}^{R} \Delta_r = M \tag{6.10}$$

$$\Delta_r \geq 0 \quad \forall r, \quad r = H(k) \tag{6.11}$$

where the mapping function $H(k)$ uniquely maps a user index k to a relay r. The solution to the problem requires solving : resource allocation variables x_{km} and bandwidth allocation variables Δ_r. We note that (6.6)-(6.11) can not be solved iteratively, by treating $\Delta_r \forall r$ as constants in iteration n and then by treating $x_{km} \quad \forall m, \forall k$ as constants in iteration $n+1$. The reasoning is as follows. In iteration n, let variables x_{km} are solved by treating Δ_r as constants. Assume that for a user k' in relay cell r', the data rate inequality in (6.7) is met with equality. In iteration $n+1$ because of the constraint in (6.7) for user k', $\Delta_{r'}$ cannot be increased further. Thus, for all the other users in cell r', there is no benefit from iteration $n+1$.

To solve the problem efficiently, our idea is to decouple the two problems. Importantly, we note that if the data rates are feasible in the r^{th} relay's access link (to its users), there exists a corresponding solution $\Delta_r > 0$ for which it is also feasible in the end-to-end link. The converse part is also true: if the data rates are not feasible in the relay's access link, there does not exist a solution $\Delta_r > 0$. Based on this fact, we use the notion of feasible list to decouple the problem. A feasible list F_r is defined as the subset of users in a relay cell r for which a scheduling solution for data rates $\Omega_k, \forall k \in F_r$ exists in the access link. Thus, $LT \sum_m x_{km} u_{km} \geq \Omega_k, \forall k \in F_r$. Now to further take into account the constraint in (6.10), we decouple the problem in our proposed GBRS scheme using two sub-routines:

- Solve x_{km}, $\forall k \in F_r$, $\forall r$, $\forall m$ with an objective to minimise the feeder bandwidth Δ_r for relay r. This is called the relay resource allocation (RRA) sub-routine.
- Select the best set of feasible lists across all the relay cells $\{1, 2, 3..R\}$ such that the sum of feeder bandwidths does not exceed M in (6.10). This is called the group selection (GS) sub-routine.

In what follows, we describe the solutions for the two sub-routines.

6.3.1 RRA subroutine : solving for the access allocation variables

In the relay resource allocation subroutine, the resource variables x_{km} of the access link are solved for a given feasible list F_r with an objective to minimise the required feeder bandwidth. Implicitly, we thus set $x_{km} = 0$, $\forall k \notin F_r, \forall m$. The problem can be seen as minimising the maximum of the feeder bandwidths needed by the users in F_r. Thus we deduce for each user $k \in F_r$, the feeder bandwidth Δ_r^k required to realise an end-to-end rate of Ω_k. This bandwidth Δ_r^k is the critical bandwidth i.e, the bandwidth required to include user k in the feasible set F_r. Equating $LT\epsilon_r \sum_m u_{km}x_{km} = \Omega_k \forall k \in K_r$, and using ϵ_r from (6.5),

$$\Delta_r^k = \frac{\sum_m \sum_{k \in F_r} x_{km} u_{km} \Omega_k}{\gamma_r [LT \sum_m x_{km} u_{km} - \Omega_k]}. \tag{6.12}$$

To minimise the feeder bandwidth needed by the set F_r, we are required to minimise the maximum taken over the set of users in list F_r for each relay r as

$$\min_{x_{km}} \max_{k \in F_r} \Delta_r^k \tag{6.13}$$

From the above, we may now represent the minimum bandwidth problem for the relay cell r and given a list F_r simply as

$$\min_{x_{km}} \max_{k \in F_r} \frac{\sum_m \sum_{k \in F_r} x_{km} u_{km} \Omega_k}{\gamma_r [LT \sum_m x_{km} u_{km} - \Omega_k]} \tag{6.14}$$

$$\text{s.t }, \sum_{k \in F_r} x_{km} = 1 \quad \forall m \tag{6.15}$$

$$0 \leq x_{km} \leq 1 \quad \forall k, \forall m, \tag{6.16}$$

or equivalently

$$\max_{x_{km}} \min_{k \in F_r} \frac{\gamma_r [LT \Omega_k^{-1} \sum_m x_{km} u_{km} - 1]}{\sum_m \sum_{k \in F_r} x_{km} u_{km}} \tag{6.17}$$

$$\text{s.t }, \sum_{k \in F_r} x_{km} = 1 \quad \forall m, \tag{6.18}$$

$$0 \leq x_{km} \leq 1 \quad \forall k, \forall m. \tag{6.19}$$

6.3.2 Optimal solution

This type of problem as in (6.17)-(6.19) is a fractional linear program [113] in variables x_{km}. The optimal solution is obtained by using substitutions $z = [\sum_m \sum_{k \in F_r} x_{km} u_{km}]^{-1}$ and $y_{km} = x_{km} z$, which basically transforms the fractional problem into a standard linear program as follows by using t as a dummy variable :

$$\max t \tag{6.20}$$

$$\text{s. t }, \gamma_r [LT \Omega_k^{-1} \sum_m y_{km} u_{km} - z] \geq t \quad \forall k \in F_r, \tag{6.21}$$

$$\sum_{k \in F_r} y_{km} - z = 0 \quad \forall m \tag{6.22}$$

$$0 \leq y_{km} \leq z \quad \forall k, \forall m, \tag{6.23}$$

$$\sum_m \sum_{k \in F_r} y_{km} u_{km} = 1. \tag{6.24}$$

(6.22) and (6.23) are written from (6.18) and (6.19) respectively. (6.24) appears because of the substitutions. The solution for (6.21)-(6.24) can be obtained using linear programming solvers such as LIPSOL. However, it can be computationally expensive for real-time implementation.

6.3.3 Sub-optimal method

We now present a simplified algorithm for resource allocation that can be implemented with relative ease at r^{th} relay. This scheme is called Heuristic-GBRS.

Step 1 : First, we ascertain if an admission list $A_r \subset K_r$ is a feasible list. The requirement is to obtain a feasibility certificate via scheduling but with low complexity. For this we perform an external point descent as in LBORA scheme in Section 4.3.4.4 as follows.

Step 1.1 : Sum rate maximisation: Greedy resource allocation is performed to allocate the resource blocks to users with highest spectral efficiency. On a per resource block basis, this is done as

$$k^* = \arg\max_{k} u_{km}, \quad \forall m \tag{6.25}$$

From (6.25) we basically obtain initial binary solutions of x_{km} as $x_{k^*m} = 1 \quad \forall m$ and $x_{km} = 0 \quad \forall k \neq k^*, \forall m$.

Step 1.2 : Two user sets G_1 and G_2 are formed wherein : G_1 is the set of users for who the solutions x_{km} in Step 1.1 satisfy the targeted data rates and G_2, the set of users for who the solutions does not provide the data rate. Thus some resource blocks are de-allocated from users in G_1 and allocated to G_2.

Step 1.3 : Resource block reallocation: A set of resource blocks V are pooled and is considered transferable from G_1 to G_2. The pool V is formed such that none of the bit rate targets of the user set G_1 would be sacrificed if any one resource block in the pool was removed.

Step 1.4 : User selection : A user is prioritised based on

$$\tilde{k} = \arg\min_{k \in G_2} \frac{\Omega_k - \sum_m u_{km} x_{km}}{\sum_{m \in V} u_{km}} \tag{6.26}$$

The metric in (6.26) is the same as (4.43) which exploits user diversity by doing user selection on the basis of average channel quality and the rate demand. To benefit from frequency diversity, we assign a resource block to $\tilde{k}$ on the basis of spectral efficiency as

$$m^* = \arg\max_{m \in V} u_{\tilde{k}m}. \tag{6.27}$$

Update $x_{\tilde{k}m^*} = 1$, and $x_{km^*} = 0 \quad \forall k \neq \tilde{k}$. The resource block m^* is removed from the set V. Steps 2) to 5) are repeated until: a) all users are given their bit rates, or b) the resource block set V is empty. Compute the total allocated data rates to the list A_r as $\sum_m u_{km}, \quad \forall k \in K_r$.

Step 2 : If the list A_r was found to be infeasible, this list is ignored, we go back to Step 1 and consider another admission list. If the admission list is feasible, we proceed further. In this case, we compute the feeder bandwidth required for each user in A_r based on (6.12). The computed feeder bandwidths is stored in a list 'BW'.

Step 3 : We now employ scheduling once more to minimise the maximum bandwidth in BW. To do this, resource blocks are reassigned from a user k_{min}, who requires the least bandwidth in the list BW to a user k_{max} who requires the highest bandwidth. The pool of resource blocks allocated with k_{min} is denoted V_{min}. To exploit frequency diversity, we select the best resource block using $m^* = \arg\max_{m \in V_{min}} u_{k_{max}m}$. Further, for a more precise computation, the resource block m^* is subdivided into g granular blocks and optimum number of granular blocks are reassigned from m^*. Steps 2, 3 are repeated until $\max BW - \min BW < \delta$, where δ is sufficiently low.

6.3.4 Group selection subroutine : solving for the feeder bandwidth variables

In this subroutine, the solution of feeder bandwidth to each relay is obtained along with finding the best set of feasible lists across the relays. Let us denote by F_r^i, the i^{th} feasible list in a relay cell r. Now the best subset of lists in all the R relay cells , out of all possible lists $F_r^i, \forall i, \forall r$ has to be found. Let the feeder bandwidth solution of the RRA subroutine to a list F_r^i be $\Delta^{min}(i, r)$. We use binary variables s_{ir} to indicate the choice of F_r^i. We now solve the following objective

$$\max \sum_i \sum_r s_{ir} |F_r^i| \quad s.t \tag{6.28}$$

$$\sum_i \sum_r s_{ir} \Delta^{min}(i,r) \leq M \tag{6.29}$$

$$s_{ir} \in \{0,1\}\,. \tag{6.30}$$

(6.28)-(6.30) is an integer program for which optimal solutions can be found by standard techniques. We however propose to use a suboptimal gradient scheme. In each step of the ascent, we merely select the 'user list-relay' pair (i^*, r^*) with the minimum gradient $\frac{\Delta^{min}(i,r)}{|F_r^i|}$, $\forall i, \forall r$ and set $s_{i^*r^*} = 1$. The algorithm terminates when $\sum_i \sum_r s_{ir} \Delta^{min}(i,r) = M$ or if all the lists have been exhausted. Upon finding $s_{ir}, \forall i, \forall r$, the feeder link bandwidth to a relay r is obtained as

$$\Delta_r = \sum_i s_{ir} \Delta^{min}(i,r), \tag{6.31}$$

Protocol overhead

The previous sections provided solutions to solve variables x_{km} and Δ_r given a set of feasible lists. However, we recall that in a relay cell r, there are K_r users. Thus there are $\sum_{i=1}^{i=K_r} C(K_r, i)$ number of user lists that are possible in each cell, where $C(n,x) = \frac{n!}{x!(n-x)!}$. For each of these lists, the RRA subroutine has to be implemented and the feeder bandwidth has to be informed to the base station. Conveying this information to the base station can be bandwidth expensive.

To reduce complexity and the signaling overhead, we propose to sort the users using an admission control metric $\frac{\sum_m u_{km}}{\Omega_k}$ at each relay. Here u_{km} is the MCS value in terms of bits per subcarrier and Ω_k bits is the demand from the user. This ordered user list is called $LIST$. The first i users of $LIST$ in relay cell r make the i^{th} admission list, and if it is a feasible list, the same is denoted as F_r^i. Note that there are k_r users and thus $i = 1..k_r$.

This effectively means that each relay cell now only needs to feedback K_r values of feeder bandwidth request , i.e, inform $\Delta^{min}(i,r)$, for $i = \{1,2,3..k_r\}$. The base station applies the group selection subroutine and decides the best value $i = i^*$ for each relay. This is in fact a hierarchical implementation of admission control : each relay makes K_r admission lists and informs the corresponding K_r feeder bandwidth requests to the base station while the base station decides the best admission list out of the K_r lists of each relay. Finally, from the solutions x_{km}, Δ_r and F_r^{i*}, we simply back substitute in (6.5) and obtain the optimal duration ϵ_r for relay r.

6.4 Comparison to other LTE relay scheme

For comparison to other forms of DF relaying in LTE, we refer to the scheme in [91] which uses a synchronised duplex time-sharing protocol.This means all the relays are activated synchronously and for the same duration. Following the approach in [91], each feeder link is active for a time duration ρ_r using bull bandwidth and all the relays concurrently transmit for a time duration α. Thus $\sum_R \rho_r + \alpha = 1$. If the relays employ round robin scheduling, the optimal duration for which all the relays are 'on' is given as

$$\epsilon(R) = [1 + \sum_{r=1}^{r=R} \frac{\sum_{k \in K_r} \sum_m u_{km}}{k_r M \gamma_r}]^{-1}, \tag{6.32}$$

In applying the scheme of (6.32), we have assumed that each relay employs round-robin TDMA/FDMA as channel access method on the access link. This is a fair comparison because in both the schemes, the same downlink bandwidth as well as frequency reuse is used at the relays. Moreover to incorporate admission control into their scheme, we follow the principle of [91] and approach it as the selection of the number of relays. The necessity of admission control is to prevent overload when the demand is high.

Thus in the TDMA/FDMA scheme the following simple admission control scheme is used at base station. The relays are enumerated in a random order by the base station and only as many relays that will maximise the number of guaranteed bit rates in (6.3) are admitted. The differences between the TDMA/FDMA scheme using [91] and our work presented in Section 6.3 are the following

- independent duplex sharing at each relay i.e, asynchronous relays
- an improved resource scheduling scheme at relays
- a novel hierarchical admission control approach at relay and base station

6.5 Results

6.5.1 Numerical example

Example :

	U1-U2	U3-U4	U5-U6	U7-U8	U9	U10-U12
RBs with MCS 9	1-11	12-22	23-33	34-43	44-48	none
RBs with MCS 6	12-37	1-11,23-37	1-22	11-33,44-48	1-38	1-48
RBs with MCS 3	38-48	38-48	38-48	1-10	39-43	none

Table 6.1: Example of modulation and coding values in resource blocks for different users. The resource block indices corresponding to the MCS levels 9, 6, 3 are shown.

Consider a relaying scenario with 12 users per relay and 4 relays. Let each user demand 2 Mbps data rate. Assume a 30% downlink overhead loss, so the overall link target is 2.8 Mbps per user. This data rate can be realised by loading 1400 bits in 0.5 ms. For example, the target can be met by allocating one resource block (which has 144 data symbols, refer the definition in Section 4.1) to each user and by using a MCS of 64 QAM R$\frac{5}{6}$ per spatial stream. This corresponds to 10 bits per subcarrier when there are two streams.

Assume a MCS distribution as in Table 6.1 with variation from 3 to 9 bits per subcarrier in the frequency domain. The abovementioned two schemes work as follows. Each of the 4 relays has the same distribution.

In the TDMA/FDMA scheme, 4 resource blocks are allocated to each of the users $U1...U12$ which exhausts all the resource blocks. Thus with equal partitioning in a TDMA/FDMA manner, each user gets an average of 6 bits per subcarrier. Applying the optimisation of [91] in (6.32), the half duplex time-sharing values ϵ for 1,2,3 or 4 activated relays are deduced to be 0.6250, 0.4545, 0.3571, 0.2778 respectively. It can be easily verified, that the maximum number of supported users occurs when only 2 relays are 'on', and thus optimum choice of $\epsilon = 0.4545$. Based on this scheme, a total of 24 users are supported with an effective loading of $6 \times 4 \times 144 \times 0.4545 = 1570$ bits per user in a time slot.

Now we illustrate our proposed scheme with the same example. Consider a user admission list

	U1	U2	U3	U4	U5	U6	U7	U8	U9
Allocated RBs	1-5	6-10	12-16	17-21	23-27	28-32	34-38	39-43	44-48
Shared RBs	11	11	22	22	33	33	-	-	-

Table 6.2: Resource block allocation in RRA subroutine. Each user gets equal data rate.

of $U1...U9$. Users 10, 11 and 12 are relatively inefficient in peak channel conditions. From Step 1 in Section 6.3.3 , a feasibility certificate for this user list can be easily obtained in the access link. The feasibility certificate basically finds a resource block scheduling that can load 1400 bit per user in a 0.5 ms time slot from relay to user.

After this a suboptimal resource block allocation is again performed in RRA subroutine (as in Step 3, Section 6.3.3) for users $U1..U9$. The RRA subroutine results in an allocation as shown in Table 6.2. In this case, each user gets an equal share of the best overlapping resource blocks (corresponding to MCS value 9), which are referred to as 'shared RBs' in Table 6.2.

As a result, 5.5 resource blocks with efficiency of 9 bits per subcarrier (MCS value 9) are allocated per user in the list $U1...U6$. 5 resource blocks of MCS value 9 are given to each of the users $U7, U8, U9$. Applying these values to (6.12), the value of $\Delta_r^k = \frac{48\times1400\times9}{10\times(6480-1440)} = 12, \quad k = 7, 8, 9 \quad \forall r$. It is noted that the corresponding duplex sharing value $\epsilon_r = 0.216 \quad \forall r$. The end-to-end effective loading is thus $9\times5\times144\times0.216 = 1400$ bits. This means that we can allocate 12 resource blocks in the feeder link to each of the 4 relays to obtain 9×4=36 satisfied users. This example aptly shows that joint resource allocation and admission control at a relay is efficient for the overall network.

In reality, we may have different MCS across the users because of different user positions in a relay cell. Quite clearly the number of supported users may vary with the user positions. To capture the expected performance in a cell, we take a statistical average by randomising the user positions.

Table 6.3: Downlink System and Simulation Parameters

BS transmit power : 43 dBm	Macrocell Bandwidth : 20 MHz
BS downtilt = 10°	Relay power control : off
Half duplex relay = yes	User scheduling at relay = yes
Simulation variables : K_a, R	PER : 10^{-2}
Relay transmit power : 23 dBm	Bandwidth per relay : 20 MHz
BS-to-RN : 2×2 MIMO	RN-to-UE MCS : per resource block
UE-to-RN : 2×2 MIMO	Carrier frequency : 2.6 GHz

6.5.2 Simulation based results

6.5.2.1 Setup and Parameters

We evaluate the performance of the proposed guaranteed bit rate by relay scheduling (GBRS) scheme numerically using channels from indoor measurements that were reported in [118]. Two sets of measurements were performed : one from an indoor relay and another from an outdoor base station. The indoor relay consists of two units : feeder unit and access unit. The two units are spatially separated but interconnected using an ethernet cable. The distance between the macro-base station and feeder unit is approximately 500 m. The transmission parameters are shown in Table 6.3

The key observations in the measurements are summarised below. In direct downlink, the signal to noise ratio (SNR) averaged over all the spatial locations was 27.5 dB. In a few locations, the SNR is as high as 40 dB and in some locations only 3 dB. A relay feeder unit was installed at a specific spot, close to the window where the averaged SNR over frequency subcarriers was 40 dB for the feeder link. We assume this will be done in practice. The relay access unit was placed indoors, few meters away from the feeder unit. The SNR variation at the user equipment indoors was from 15 dB to 60 dB in spatial location which averages to 57 dB. Interestingly, in both the macro-link and relay link to the user, the locations of low SNR were identical and occurred at deeply shadowed locations indoors.

Feeder link : The SNR required for loading 64 QAM R$\frac{5}{6}$ (5 bits per subcarrier per stream) on the feeder link with a target packet error rate (PER) of 10^{-6} is 21.5 dB. Note that the average

SNR was 40 dB at the feeder unit of relay. We can therefore work on a 40 dB -21.5 dB=19.5 dB link budget margin on the feeder link. Therefore a base station can employ 19.5 dB power control on feeder link. It is seen that upon employing a transmit control of 19.5 dB, the base station signal power is more or less expected to be along the noise floor in most indoor locations. Figure 6.4 shows the received signal power at user equipment, comparing from the base station after 19.5 dB power control to the reception from relay. As can be seen, the relay signal power is overwhelmingly higher in all locations. Thus in our simulations, we simply ignore any interference from the feeder link of one relay to another relay's access link. We remark that for more efficiency, the base station may load the data on selected resource blocks in the feeder link to take advantage of the frequency diversity. It can be further noted that if the PER target is 10^{-2} the SNR can be reduced to 14.5 dB. Thus, there is an extra 7 dB link budget margin for the feeder link, in case the signal loss is unexpectedly higher or more even higher power control.

Access link : Each user equipment reports the best precoding matrix indicator (PMI) to the transmitter for each resource block. There are 144 data symbols in a resource block as before. The received SNR on a subcarrier for the two spatial streams is estimated using a MMSE receiver. From the SNR, the MCS level per spatial stream is selected from among quantised MCS levels, with a target PER options of 10^{-2} or 10^{-6}. The lowest modulation is QPSK and the highest is 64-QAM with coding options according to the MCS table. Information about the best PMI and MCS is conveyed to the transmitter according to a feedback scheme.

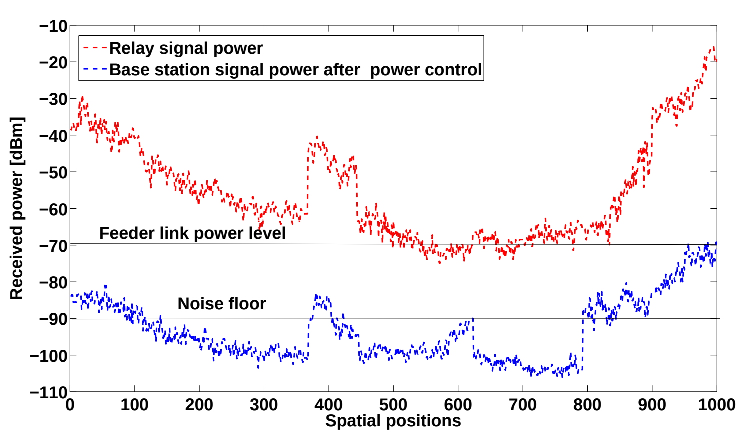


Figure 6.4: Results showing suppression of interference in indoor locations from the base station

For numerical evaluations, we assume R indoor cells, each identical to our measurement site and provided with one relay each. The feeder link capacity is 10 bits per subcarrier. In one channel realisation, K_a users are dropped randomly in each indoor cell and every user demands a bit rate of 2 Mbps. Thus there are totally $K_a R$ users per realisation. Through this, we can vary a) the number of users per relay cell $K_r = K_a, \forall r$, and b) the number of relay cells R. The results show a statistical average of the number of supported users by randomising 2000 channel realisations. The following schemes are studied:

a) optimal relay resource allocation and group selection, as in Sections 6.3.2 and 6.3.4

b) suboptimal relay resource allocation and group selection as in our proposed Heuristic-GBRS, Sections 6.3.3 and 6.3.4

c) 'macro-downlink using LBORA', i.e, scheduling user and frequency selection (without any power control , without relay) with the LBORA scheme in Section 4.3.4.4 , Steps 1.1-1.4

d) TDMA/FDMA relaying scheme of [91] for comparison in Section 6.4.

We study four scenarios as follows in Table 6.4.

Table 6.4: Simulation test scenarios

	Users/relay	Num.relays	PER	Feedback	Expected coverage
Scenario A	1 to 12	4	10^{-2}	26 MCS set + precoding per RB	Somewhat optimistic
Scenario B	4	1 to 12	10^{-2}	26 MCS set + precoding per RB	Most optimistic
Scenario C	1 to 12	4	10^{-6}	Fine feedback , refer Table 5.1	Very limited
Scenario D	4	1 to 12	10^{-6}	Fine feedback , refer Table 5.1	Somewhat limited

Scenarios A and B use the same reference MCS set for PER 10^{-2} as in WINNER [109], .

Scenarios C and D are very low PER and use the fine feedback scheme in Table 5.1, Section 5.4. Applications using TCP demand very low PER thus making these two scenarios limited in coverage.

Furthermore, with respect to relaying, we expect scenarios B and D to benefit from high number relays with low user density in each relay. The plots are attached at the end of the chapter.

6.5.2.2 Comparison in optimistic coverage scenarios

We first compare the optimistic coverage scenario of PER target of only 10^{-2} with and without relaying. This corresponds to scenarios A and B.

Scenario A: Dense relay cells

In Figure 6.5a we show the number of satisfied users against the number of active users (contending users) per relay cell for a PER target of 10^{-2}. The number of relay cells is fixed to 4 while the number of users per relay is varied from 1 to 12. It is clear that when the number of users increase, the user density within the relay cell increases and thus we call it as dense relay cells.

As can be seen from Figure 6.5a, 'macro-downlink with LBORA' supports approximately 32 users across the 4 relay cells. This schemes exploits multiuser diversity by means of frequency diversity and user selection in the macrocell. Therefore, a base station transmitting with maximum signal power of 43 dBm can use this scheme to realise reasonable coverage. The 'red-circle' curve shows the result for round-robin TDMA at relays and with relay admission control at the base station. This plot clearly shows that performance of TDMA DF relaying suffers in dense relay cells. In fact, the TDMA relay with admission control comes worse off than even the direct downlink with only 22.5 satisfied users (on average) when there are 12 users in each relay. One main reason is that a TDMA half duplex relay sacrifices degrees of freedom twice: for half duplex time sharing and then user time sharing. Relay admission control, which is an incorporated feature in the scheme, provides some robustness for high number of contending users but not a marked improvement.

Using our proposed GBRS scheme, we are able to achieve a substantially higher number of satisfied users with the same relay set up. The proposed scheme makes use of dynamic channel allocation to maximise the performance according to the channel states. On an average 37.5 users are supported with the suboptimal heuristic-GBRS scheme. The heuristic is a fast allocation which is suitable for real-time implementation. The optimal allocation achieves 38.5 users. This represents 18 % improvement as compared to direct downlink.

Scenario B: Lightly loaded relay cells

We now fix the number of users per relay to only 4 but increase the number of relays across the macro-cell. Note that for a fair comparison between scenarios A and B, the same feeder link bandwidth of 20 MHz is employed and the same PER target 10^{-2} is applied. This means that this scenario is constrained by the feeder link bandwidth which does not scale up with the number of relays. Figure 6.5b shows the results. This scenario is less restrictive on TDMA DF relaying which can now support up to 30 users. Thus TDMA/FDMA relaying can support more users in Scenario B as compared to Scenario A. One reason is because we now have only 4 users per relay cell, and the role of relay admission control becomes more effective. The direct macro-downlink obviously performs identical to Scenario A because there are no relays involved in the macro-downlink case and hence the two scenarios are identical. Our proposed heuristic-GBRS scheme provides benefit also in lightly loaded cells and supports 43 users as compared to direct downlink's 32, which is a 34 % improvement for this scenario. With less number of users per relay, the duplex time sharing ratios become small at each relay. This benefit is passed on to the feeder link as requiring less bandwidth.

6.5.2.3 Comparison in limited coverage scenario

In these comparisons, we now look at relaying performance in Scenarios C and D. In these two scenarios, the link QoS targets have become stricter. A PER of 10^{-6} is required and only limited feedback, as in the fine feedback scheme of Table 5.1 can be sent on the uplink. Thus these are limited coverage scenarios.

Scenario C: Dense relay cells In Figure 6.6a, we look at dense relay cells where there can be a large number of users per relay cell and a low PER target is applied. Thus the congestion per relay cell is high while the link coverage is also limited, making it the worst performance scenario.

To observe the effect of link QoS targets on achievable coverage in terms of number of users, this scenario C can be compared to Scenario A.

We straightaway see that macrocell with LBORA (without relay) can support only 24 users as compared to 32 previously in Scenario A. This shows a drop of 25 % in number of users when the link targets become stricter.

Our proposed relaying heuristic-GBRS scheme can support up to 36 users. Remarkably, this is almost the same as Scenario A, where it was 37.5 users. Thus we see that our relaying solution can still provide excellent coverage even though the link QoS constraints are strict. Overall, there is a 50% increase in number of users from 24 to 36 as compared to the direct macrodownlink.

The comparative TDMA/FDMA relaying with admission control can support only 18.5 users on average. This again shows a drop of 18 % in number of users as compared to Scenario A for the same scheme. Thus, this scheme comes off as the worst performing scheme in this scenario.

Scenario D: Lightly loaded relay cells

This is shown in Figure 6.6b for a PER target of 10^{-6}. As before, our proposed schemes heuristic-GBRS and optimal-GBRS are again able to provide substantial coverage benefit. The number of users is 42, which is very close to the result in Scenario B. Thus, the link QoS constraints in terms of low PER and low feedback have no major effect on relaying performance.

Direct macrocell and TDMA/FDMA relaying both under perform with respect to our proposed relaying scheme. In this case, the TDMA relaying scheme supports 24 users, thus slightly benefiting from the low load in relay cell. The direct downlink becomes the worst performing scheme in this case with less than 24 users.

Finally, we remark that although we have not shown in the results, the benefits of 19.5 dB base station power control on relay feeder link can be passed on as interference reduction to other relay cells. This insight can be incorporated in system level results.

6.6 Conclusion

High data rate demand by indoor wireless users can be expected in future systems. The mean spectral efficiency of outdoor to indoor communication is observed to be less than 2 bits/sec/Hz. Therefore a large amount of system bandwidth may be consumed by indoor users.

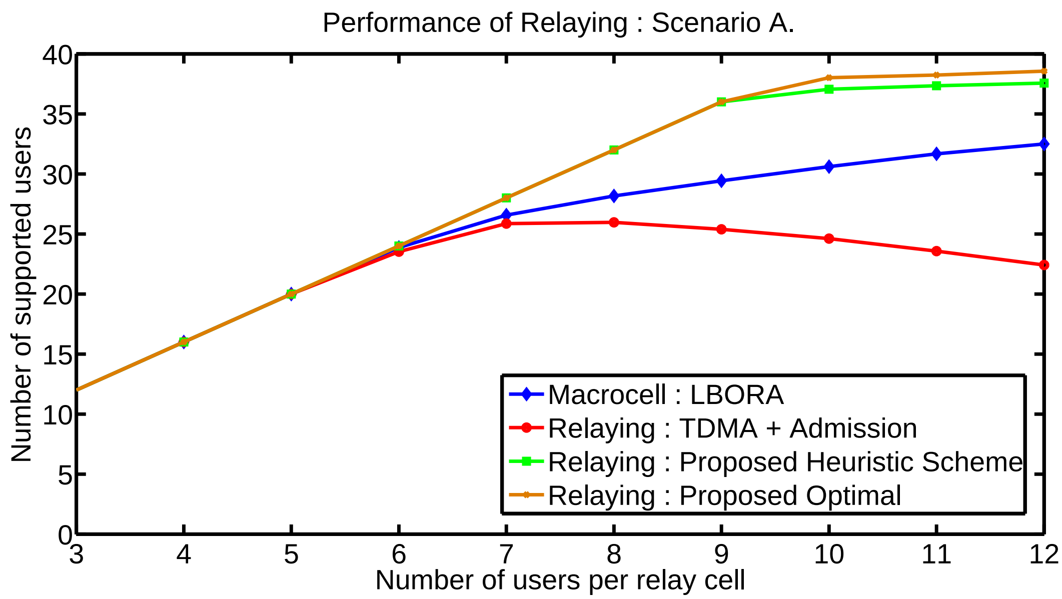
A convenient way of addressing this issue is indoor wireless relaying. In this chapter, we have introduced a new scheme called guaranteed bit rates by relay scheduling (GBRS), in order to extend the coverage of GBR users.

The deployment idea is to let each relay employ frequency reuse and further activate many such relays by exploiting spatial separation and building penetration losses. We employ hierarchical OFDMA, wherein multiple relays and multiple users per relay are served in a dedicated time slot. The dedicated time slot is split into two phases in time division. In the first phase, the feeder link to relay is active. In the second phase, the access link from the relay to users is enabled. The second phase allows each relay to utilise scheduling degree of freedom by using OFDMA transmission to users.

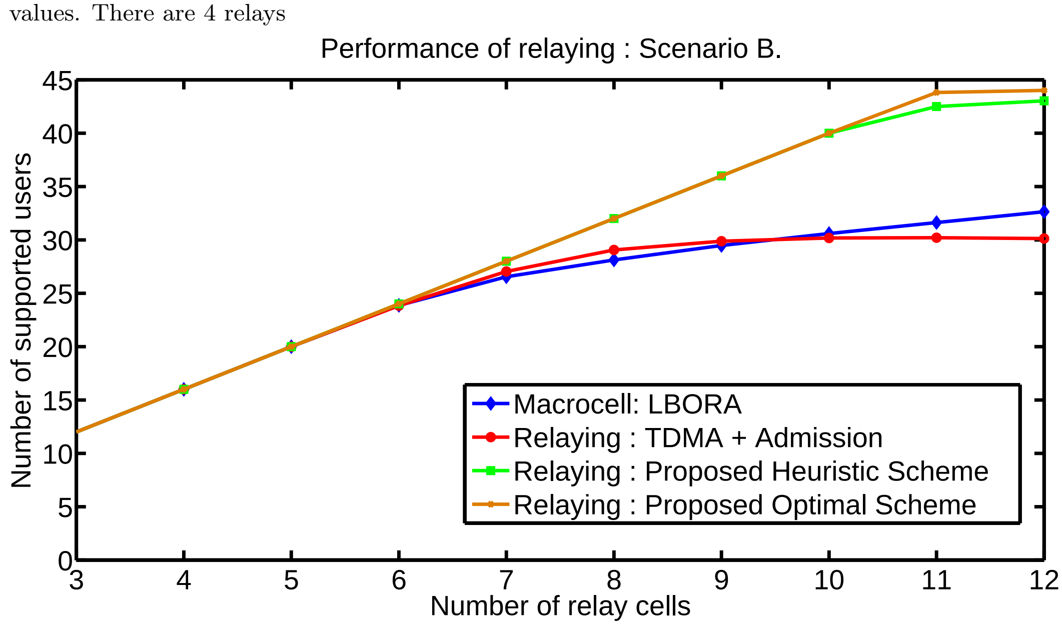
The problem is defined mathematically for optimal usage of the bandwidth. This problem turns out to be hard to solve and can not be done in iterations. We then break down the problem in to two sub-problems of relay resource allocation (RRA) and group selection (GS). RRA performs scheduling at relays to minimise the feeder bandwidth request to the base station, and GS lets the base station perform selection of relays and sub-groups of users on each relay. In effect, RRA reduces the active time of relays in the access link. Optimal and heuristic solutions are provided.

Upon relaxation of variables, the optimisation of RRA becomes a fractional linear program. The resulting solutions for resource block allocation are continuous. The heuristic solution called as heuristic-GBRS is a fast integer solution and results in allocations with a fixed granularity. The heuristic is more suitable for real-time implementation as it involves only sorting operations. No large matrix inversions and step size searches are required as in the optimal solution. The heuristic also performs within 1% of the optimal.

Finally, we compare the performance of the GBRS scheme to a fixed TDMA/FDMA resource allocation at relays in four test scenarios. All the four scenarios target 2 Mbps data rate per user. Surprisingly, we find that the TDMA/FDMA relay scheme is beneficial in only one scenario. It may even perform worse than direct downlink (without a relay) in other 3 scenarios. This is on account of the capacity penalty incurred by duplex time sharing between feeder link and access link. We thus conclude that TDMA/FDMA relays may not automatically bring benefits to a network in terms of the number of users. In the most favorable case, TDMA/FDMA relaying support 25 users and outperforms the macro-downlink which has no relay. Our proposed GBRS relaying scheme is suited for all the scenarios. It can support up to 42 users out of 48 contending users, with a guaranteed data rate of 2 Mbps to each user. Based on these results, we expect relays to play an important role in delivering high data rates to indoor users in future wireless networks.

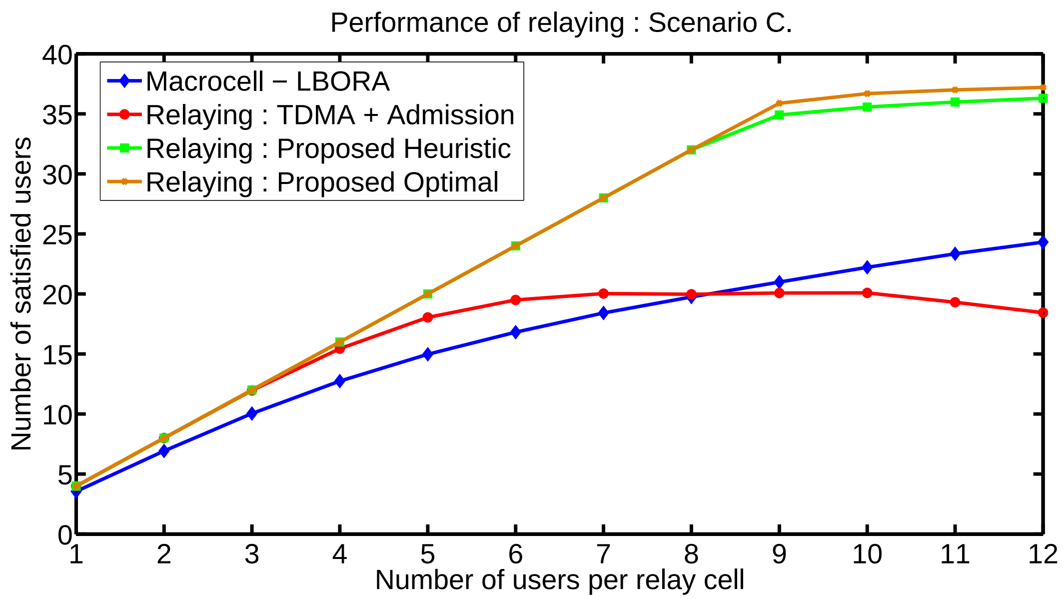


(a) Scenario A. Dense relay cell along with low link QoS constraints. Up to 12 users per relay, high PER of 10^{-2} and 26 MCS values. There are 4 relays

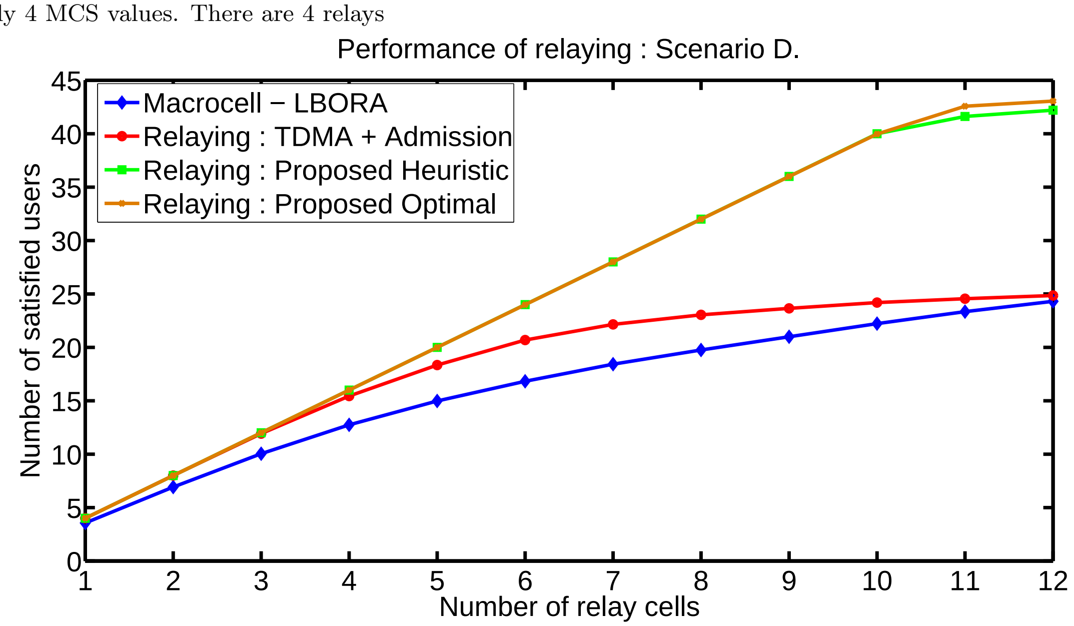


(b) Scenario B. Thin relay cells along with low link Qos constraints. It has only 4 users per relay, high PER of 10^{-2} and 26 MCS values. There are 4 users per relay

Figure 6.5: Performance of proposed GBRS relaying scheme in dense relay cells for different QoS constraints



(a) Scenario C. Dense relay cell along with high link QoS constraints. Up to 12 users per relay but low PER of 10^{-6} and only 4 MCS values. There are 4 relays



(b) Scenario D. Thin relay cells along with high link QoS constraints. It has only 4 users per relay but low PER of 10^{-6} and only 4 MCS values. There are 4 users per relay

Figure 6.6: Performance of proposed GBRS relaying scheme in lightly loaded relay cells for different QoS constraints

Chapter 7

Conclusion

7.1 Summary

The capabilities of cellular systems have been fast growing in the last decade. The latest addition to the cellular systems is OFDMA. OFDMA transmission is well suited for the future vision of 4G as a user-centric system. With OFDMA, flexible downlink resource allocations are possible depending on feedback from the user equipment. This can make cellular networks even more user-friendly because the users can now decide their service requirements.

Providing consistent data rate no matter how many users are in the system remains a particular challenge in wireless networks. An even more challenging task is to provision these data rates no matter where the users are located. This thesis follows these perspectives and considers the issue of maximising the number of users at a guaranteed data rate in a LTE-like systems. This problem is addressed in Chapters 4 and 6. Chapter 4 provides OFDMA resource allocation for direct macro-downlink with a scheme called Load Balanced Opportunistic Resource Allocation (LBORA), while Chapter 6 provides OFDMA resource allocation with relays using a scheme called Guaranteed Bit Rate by Relay Scheduling (GBRS).

To show our results we have used channel values obtained from measurements throughout this thesis. Some real-time data rate measurements are also shown in Chapter 5. A 3GPP LTE prototype user equipment and LTE-like base station at Heinrich-Hertz Institute in Berlin were used to collect the measurements. Numerical results showed that the proposed scheduling method LBORA provides excellent gains in the number of guaranteed bit rate users in a macro-cell. The novel proposed GBRS scheme is able to extend the performance of LBORA by including relays which caters for users in deeply faded locations.

In Chapter 4, OFDMA resource allocation for two service types are discussed : non-guaranteed and guaranteed bit rates. Proportional fair and utility proportional fair allocation methods are shown for non-guaranteed bit rates.

Guaranteed bit rates are considered in Section 4.3 in Chapter 4. In this case, we formulated the problem of maximising the number of users at a given data rate. It was shown that the problem can be solved via a linear program. Yet, linear programs with a large number of variables are difficult to implement in real-time. Thus in Section 4.3.4, a heuristic scheme called 'Load Balanced Opportunistic Resource Allocation' was presented. The key idea of the scheme is to integrate admission control with resource allocation, which we call joint admission control and resource allocation. The LBORA scheme is further improvised using a modification of the Hungarian algorithm, known as the k-cardinality algorithm. Performance of the LBORA scheme is compared to other scheduling techniques. Results based on simulations showed a gain of 58% and 78% in the number of users for data rates of 1.45 Mbps and 5.8 Mbps respectively. A running numerical example is provided throughout the chapter for the sake of illustration.

Even though scheduling can support a large number of users, coverage issues may still arise if there are coverage holes in a cell. Coverage holes are areas in a macro-cell with significantly lower received power. One such key scenario is outdoor to indoor coverage because a large amount of traffic can generate from indoors but building penetration losses can degrade indoor reception. For solving indoor coverage holes, a deployment concept using decode-and-forward relay is proposed. The basic idea is to use the full bandwidth (20 MHz) on the access link. We make use of the building penetration loss to isolate interference from the other cells.

To investigate the coverage more thoroughly, we reported indoor coverage measurements in Chapter

5. These measurements confirmed the existence of coverage holes from an outdoor base station. Another set of measurements performed with relays confirmed that relays alleviate the problem. These real-time measurements are shown in Section 5.5 of Chapter 5.

Adaptive modulation plays an important role in the notion of coverage. We further classified three feedback schemes for adaptive modulation in OFDM named coarse, fine and extended in Section 5.4 of Chapter 5. Coverage with and without relaying are compared with respect to feedback. The measurements revealed the following :

- Indoor coverage holes do exist in macro-OFDM cells at 2.6 GHz carrier frequency.
- The indoor data rate from outdoor base station can vary sharply anywhere between 0 Mbps to 160 Mbps within 10 meters.
- Limitation in uplink feedback can severely worsen the macro-coverage without a relay.
- One indoor relay can provide full coverage to a large office floor of dimension 20 $m \times 10$ m even with a relatively low transmit power (23 dBm). The signal from the relay can also provide coverage for one adjacent floor in the same office building (3 floors in the building).

An insight that is available from our results is that there could be benefits from switching the feedback between the three levels of coarse, fine and extended for different users. In this context, a novel scheme called user selective feedback scheme is presented in Section 5.8. This scheme allocates feedback levels to users depending on their channel conditions and uplink feedback capacity. The benefit of the scheme is illustrated in a multi-user scenario with proportional fairness. When the uplink feedback is reduced by 25%, the performance loss of the relay link is only less than 5%. Given that the uplink cellular channel may have to cater for hundreds of VoIP users, we can expect this type of deployment to provide more benefits in real-time. In addition, a first insight is offered regarding the trade-off between network efficiency and user efficiency with relaying. The trade-off is dependent on the amount of feeder bandwidth. This observation is further made use of to develop the novel proposed GBRS scheme in Chapter 6, where we support multiple relays.

In Chapter 6, we have created the GBRS relaying framework to maximise the number of indoor users, termed guaranteed bit rate by relay scheduling (GBRS). In the scheme, many relay cells receive data from a serving macro-base station and then forward the data to multiple users in their respective cells. The optimisation problem is shown to be hard to solve as it involves joint OFDMA optimisation on feeder and access links. We use a novel decoupling scheme to break down the problem as *bandwidth request* from relay and a *bandwidth grant* at the base station. A heuristic called Heuristic-GBRS is proposed for a fast solution. Relay resource allocation (RRA) and group selection (GS) execute the above two tasks respectively.

Even though it may appear that relays represent additional resources such as transmit power and bandwidth, they are also constrained by time delays. Thus our relaying protocols employ a fixed amount of time, and complete the feeder and access transmission within this time duration. This constraint makes one wonder whether duplex time sharing optimisation brings large benefits. This was the main theme of the work in Chapter 6.

Coverage analysis was performed for four test scenarios. It is seen that when resources are not well managed, the advantages of relaying can vary depending on the test scenario. On the other hand, resource management proves to be key to obtain the full benefits of relaying under all scenarios. This is even though the relays advantageously exploit full frequency reuse indoors.

A case in point is TDMA/FDMA relaying with indoor full frequency reuse. The benefit of TDMA/FDMA relaying is evident in only one test scenario, where there are 6 relay cells with 4 users per relay and a required PER is 10^{-6}. The GBRS scheme outperforms TDMA/FDMA relaying in all the test scenarios. It is also superior to the direct macro-downlink in all the test scenarios, providing 100 % improvement for 10 relays and 4 users per relay.

7.2 Benefit of relays based on our scenario

Results in interference-free scenarios have shown that relaying can increase the number of users. We present a simple case for operational benefit without taking into account operational costs such as maintenance, site acquisition and other business factors. Consider that an operator has a spectrum license in an area of 9 $km \times 9$ km and installs many macro-cell sites. There are n number of macro-sites with a 20 MHz is deployed at each site. A typical value of n is 50. Assume there are M indoor cells in each site which are identified to be hotspots. In a 500 m macro-cell, there may be $M = 20$ indoor cells. Out of the M hotspots, let 10 be active at a time. 4 users are active within each indoor cell (based on some user activity factor). The users are negotiated for 2 Mbps Internet connection. Thus there are a total 40 users who expect 2 Mbps service. Our results show that the direct downlink from the base station can at most supports 20 users (with our scheduler at the base station) out of the 40 users by using on an operational bandwidth of 20 MHz. We however recall that our outdoor to indoor tests were conducted in an *in an interference-free scenario* at the cell edge at a distance of 500 m. A frequency reuse factor of r may thus be needed to realise such a possibility out of frequency planning. A conservative (an optimistic) value of r would be 2 to realise an intereference-free downlink because a reuse of 1 would clearly introduce interference. If relays are well deployed, all the 40 users can be supported. There is however the equipment cost of relays. If each relay costs x units, the total equipment cost of relays is $n \times M \times x$.

One easy alternate option for coverage without relays may be to deploy additional bandwidth for that license area. To cover the remaining 20 users per macro-cell, another 20 MHz band is needed which will cost y units. For a frequency reuse factor of r (applied on the n macro-sites), an additional spectrum of $20 \times r$ MHz is needed for that license area, costing $y \times r$ units. With this simple calculation, relaying is profitable for an operator as long as the spectrum to relay cost ratio is $\frac{y}{x} > \frac{nM}{r}$.

7.3 Future work

There are a variety of areas of future research that is possible on the topic of cellular downlink resource allocation and relays. Few of them are mentioned below.

- System level study : Our results were obtained in a single-cell scenario with multiple users in the cell. Application of these techniques in a system level scale, incorporating sectorisation and adjacent cells can be done. Specifically, the working of our scheduling technique in the case of handover may need more investigation. Much more research is necessary to obtain a better understanding of these issues.

- High mobility : Most of our measurements were conducted under conditions of low mobility up to 40 Kmph. Tests on scheduling with the proportional fair algorithm has shown that the scheduler can handle this mobility [120]. However, the same result may not true at high mobility where the channel outdate time is short. More research is needed to develop techniques which are adaptable to this environment.

- Further optimisation : Although our scheduling techniques show significant performance improvement in OFDMA, they are not optimised using transmit power allocation on subcarriers. Our optimisation solely focused on resource block allocation with uniform power allocation. This issue can be addressed in future work.

References

[1] S. Fratassi, H. Fathi, F. H. P Fitzek, R. Prasad, M. D. Katz "Defining 4G Technology from the Users Perspective," *IEEE Network Magazine,* vol. 20, issue. 1, pp. 35-41, January 2006.

[2] E. Ciana, M. D. Sanctis and M. Ruggieri "Convergence Towards 4G: A Novel View of Integration," *Journal of Wireless Personal Communications,* vol. 33, issue. 3, pp. 327-336, June 2005.

[3] http://www.3gpp.org/About-3GPP.

[4] 3GPP Long Term Evolution Standardisation Working Documents 36. 212, 36. 213, 36. 214 found at www.3gpp.org/ftp/Specs

[5] "Proposed Study Item on Evolved UTRA and UTRAN," RP-040461, TSG-RAN Meeting 26, 8-10 December 2004, Athens, Greece.

[6] Requirements for Evolved UTRA (E-UTRA), 3GPP TR 25.913 V2.0.0, Requirements for Evolved UTRA and UTRAN (Release 7), May 2005.

[7] Technical Specification Group Radio Access Network; Feasibility study for evolved Universal Terrestrial Radio Access (UTRA) 3GPP TR 25.912 V 7.2.0, June 2007. Available for download in www.3gpp.org

[8] A. Simonsson " Frequency Reuse and Intercell Interference Co-ordination in E-UTRA," *IEEE VTC 2007 Spring*, pp. 3091 - 3095, 22-25 April 2007, Dublin.

[9] Y. Wang, S. Kumar , L. Garcia et.al " Fixed Frequency Reuse for LTE-Advanced Systems in Local Area Scenarios," *IEEE VTC 2009 Spring*, pp. 1-5, 26-29 April 2009, Barcelona.

[10] L. Thiele, V. Jungnickel and T. Haustein " Interference Management for Future Cellular OFDMA Systems Using Coordinated Multi-Point Transmission," *IEICE Transactions on Communications*, vol. E93.B, Issue 12, pp. 3228-3237, 2010.

[11] IST-4-027756 WINNER II. Final assessment of relaying concepts for all CGs scenarios under consideration of related WINNER L1 and L2 protocol functions. D3.5.3 v1.0, 2007.

[12] 3GPP TR 36.913 v 8.0.0, Technical Specification Group Radio Access Network Requirements for Further Advancements for E-UTRA, LTE-Advanced (Release 8).

[13] ITU-R M. 2134, Requirements related to technical performance for IMT-Advanced radio interface(s).

[14] Technical Specification Group Services and System Aspects; Policy and charging control architecture (Rel 9), 3GPP TS 23.303 v 9.1.0.

[15] Technical Specification Group Radio Access Network; Further advancements for E-UTRA physical layer aspects (Release 9), 3GPP TR 36.814 V9.0.0.

[16] Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer procedures (Release 10), 3GPP TS 36.213 V10.2.0.

[17] Interference mitigation - Considerations and Results on Frequency Reuse, Meeting 42, TSG-RAN WG1 R1-050738, 29 August - 2 September, 2005, London, UK.

[18] Q. Spencer, L.H. Swindlehurst, M. Haardt, " Zero-Forcing Methods for Downlink Spatial Multiplexing in Multiuser MIMO Channels," *IEEE Transactions on Signal Processing*, vol. 52, no. 2, pp. 461-471, February 2004.

[19] Dual Polarized Base Station Antennas For Polarization Diversity Systems, http://www.kathrein-scala.com/tec$h_b$ulletins/DualPolarized.pdf

[20] Basic Antenna Principles for Mobile Communications, http://kathrein.pl/down/BasicAntenna.pdf

[21] G. Durgin and T.S. Rappaport "Measurements and Models for Radio Path Loss and Penetration Loss In and Around Homes and Trees at 5.85 GHz," *IEEE Transactions on Communications*, vol. 46, no. 11, November 1998.

[22] Building Penetration Measurements From Low-height Base Stations At 912, 1920, and 5990 MHz; NTIA Report 95-325. Available for download at http://www.its.bldrdoc.gov/pub/ntia-rpt/95-325/

[23] P. Nobles, D. Ashworth and F. Haltall "Propagation measurements in an indoor radio environment at 2, 5 and 17 GHz," *IEE Colloquium on High Bit Rate UHF/SHF Channel Sounders - Technology and Measurement*, 3 Dec 1993, London.

[24] A.M. Feldman, R. Serrano "Welfare economics and social choice theory,", 2nd edition, 2006, ISBN-13: 978-0387293677, Springer publications.

[25] V. Emamian, P. A. Anghel, M. Kaveh "Multi-user spatial diversity system in a shadow-fading environment," *Proc. IEEE VTC 2002*, vol. 1, pp. 573-576, September 24-28 2002.

[26] E. Telatar, "Capacity of multi-antenna gaussian channels," *European Transactions of Telecommunications,* vol. 10, no. 6, pp. 585-595, November 1999.

[27] G. Foschini and M. Gans "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless. Personal Communications,* vol. 6, no. 3, pp. 311-335, March 1998.

[28] G. Foschini, "Layered space-time architecture for wireless communications in a fading environment when using multi-element antennas," *Bell labs Tech J.,* , pp. 41-59, 1996.

[29] L. Zheng and D.N.C Tse "Diversity and Multiplexing: A fundamental trade off in multiple-antenna channels," *IEEE. Trans. Inform. Theory,* vol. 49, no. 5, pp. 1073-1095, May 2003.

[30] S. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE J. Select. Areas Commun*, vol. 16, no. 8, pp. 1451-1458, 1998.

[31] V. Tarokh, N. Seshadri and A.R Calderbank, "Space time codes for high data rate wireless communications:performance criterion and code construction," *IEEE Trans. Inform. Theory*, vol. 44, pp. 744-765, March 1998.

[32] B. Hassibi and B. Hochwald "High rate codes that are linear in space and time," *IEEE Trans.Inform.Theory,* vol. 48, no. 7, pp. 1804-1824, July 2002.

[33] A. Narula, M. Lopez, M. Trott and G. Wornell "Efficient use of side information in multiple-antenna data transmission over fading channels," *IEEE J. Selec. Areas Commun*, vol. 16, pp. 1423-1436, October 1998.

[34] R. W Heath RW, A. J Paulraj 'Switching between diversity and multiplexing in MIMO systems," *IEEE Transactions of Communications* vol. 53, pp. 962-968, 2005.

[35] T. Haustein et. al "MIMO-OFDM for Cellular Deployment - Concepts, Real-time Implementation and Measurements towards 3GPP-LTE," 15^{th} EUSIPCO 2007, pp. 1849-1853, Poznan, Poland, September 3-7 2007.

[36] S. Schiffermller, V. Jungnickel, "Practical Channel Interpolation for OFDMA," IEEE Globecom 2006, San Franscisco, USA, 27 November-2 December 2006.

[37] E. Lawrey 'Multiuser OFDM," *Proceedings of ISSPA'99*, Brisbane, Australia, pages 761-764, 22-25 August, 1999.

[38] W. Rhee and J. M. Cioffi 'Increase in Capacity of Multiuser OFDM System Using Dynamic Subchannel Allocation," *Proceedings of IEEE VTC Spring'00*, vol. 2, pp 1085-1089, May 2000.

[39] C. Y. Wong, R. Cheng, K. B. Letaief, R. D Murch 'Multiuser OFDM with Adaptive Subcarrier, Bit and Power Allocation," *IEEE Selected Area of Communication*, 17(10), pp 1747-1758, October 1999.

[40] J. Jang and K. B. Lee 'Transmit Power Adaptation for Multiuser OFDM Systems," *IEEE Selected Area of Communication*, vol. 21, pp 171-178, February 2003.

[41] R. Agrawal, R. Berry, J. Huang, V. Subramaniam "Optimal Scheduling for OFDMA Systems," *Proc. of Asilomar 2006* , October 2006.

[42] T. Bonald, "A score-based opportunistic scheduler for fading radio channels ," *Proc. European Wireless 2003*, 2003.

[43] F. P. Kelly, A. K. Maulloo and D. K. H. Tan " Rate Control in Communication Networks: Shadow Prices, Proportional Fairness, and Stability," Journal of the Operational Research Society 49, pp. 23752, 1998

[44] P. Vishwanath, D. N. C Tse and R. Laroria, "Opportunistic beamforming using dumb antennas," *IEEE Transactions on Information Theory*, vol. 48, pp. 1277-1294, 2002.

[45] J. M Holtzmann, "Asymptotic analysis of proportional fair algorithm," *Proc. of IEEE PIMRC 2001*, F-33-F36 (2001).

[46] H. Kim, Y. Han "A Proportional Fair Scheduling for Multicarrier Transmission Systems," *IEEE Communication Letters*, vol. 9, no. 3, pp. 210-212, March 2005.

[47] P. Svedman, P. Wilson, B. Ottersen, "A QoS-aware proportional fair scheduler for opportunistic OFDM," *Proc. of IEEE VTC 2004* , vol.1, pp. 558-562, 26-29 September 2004.

[48] M. Andrews, L. Zhang "Scheduling Algorithms for Multi-Carrier Wireless Data Systems," *Proc. of MOBICOM 2007* , September 2007.

[49] P. Frjdh, U. Horn, M. Kampmann, A. Nohlgren, M. Westerlund "Adaptive Streaming withing the 3GPP Packet-Switched Streaming Service," *IEEE Network*, pp. 34-40, March 2006.

[50] Z. Jiang, Y. Ge and Y. Li, "Max-Utility Wireless Resource Management for Best-Effort Traffic," *IEEE Transactions of Wireless Communications*, vol 4, no. 5, pp. 100-111, January 2005.

[51] T. Girici, C. Zhu, J. Agre and A. Ephrimedes "Practical Resource Allocation Algorithms for QoS in OFDMA-bases Wireless Systems ," *Proc. of CCNC*, pp. 897-901, 2008.

[52] S. Pal, P. Ghosh, A. R. Mazloom, S. R. Kundu and S. K. Das, "Two Phase Scheduling Algorithm for Maximizing the Number of Satisfied Users in Multi-Rate Wireless Systems," IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks, 2007. WoWMoM 2007, pp. 1-9, Espoo, Finland, 18-21 June 2007.

[53] Y. J. Zhang, K. B. Letaief, "Multiuser Adaptive Subcarrier-and-Bit Allocation With Adaptive Cell Selection for OFDM Systems," *IEEE Trans. Wireless Commns* , vol.3, pp. 1566-1575 , 2004.

[54] Z. Zhang, Y. Le, E. Chong "Opportunistic Scheduling for OFDM Systems with Fairness Constraints ," *Eurasip Journal on Wireless Communications and Networking*, Article ID 215939 (2008).

[55] H. W. Kuhn "The Hungarian Method for the assignment problem," *Naval Research Logistics Quarterly 2* (1955).

[56] H. Yin and H. Yiu "An Efficient Multiuser Loading Algorithm for OFDM-based Broadband Wireless Systems," *Proceedings of IEEE GLOBECOM' 00*, vol. 1, pp. 103-107, 2000.

[57] Y. Zhang "Solving Large-Scale Linear Programs by Interior-Point Methods under the MATLAB Environment," *Department of Mathematics and Statistics, University of Maryland Baltimore County* , Technical Report TR 96-01 , 1996.

[58] J. Chinneck "An effective polynomial-time heuristic for the minimum cardinality IIS set-covering problem," *Annals of Mathematics and Artificial Intelligence,* , vol. 17, no. 1, pp. 127144, 1996.

[59] R. Iyengar, K. Kar, B. Sikdar, "Scheduling Algorithms for Point to Multipoint Operation in IEEE 802.16 Networks," *Proc. of WiOpt 2006*, pp. 1-7, April 2006.

[60] J. Gross, "Admission Control based on OFDMA Channel Transformation," *Proc. of 10th IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM)*, 2009.

[61] T. Girici, C. Zhu, J. R. Agre, and A. Ephremides "Proportional Fair Scheduling Algorithm in OFDMA-Based Wireless Systems with QoS Constraints," Journal of Communications and Networks, vol. 12, No. 1, pp. 30-42, February 2010.

[62] S. J. Bae, J. J. Lee, B. Choi, S. Kwon and M. Y. Chung "A Resource-Estimated Call Admission Control Algorithm in 3GPP LTE System," *Proc. ICCSA 2009*, Lecture Notes in Computer Science, Volume 5593/2009, pp. 250-260, 2009.

[63] C. Koutsimanis and G. Fodor "A Dynamic Resource Allocation Scheme for Guaranteed Bit Rate Services in OFDMA Networks," *Proceedings of IEEE GLOBECOM' 08*, vol. 1, pp. 2524-2530, 2008.

[64] B. Timus, "Cost Analysis Issues in a Wireless Multihop Architecture with Fixed Relays ," *Proc. IEEE VTC Spring 2005* , vol. 5, pp. 3178- 3182 , 30 May-1 June 2005.

[65] B. T. Walkenhorst, Achieving Near-optimal MIMO Capacity in a Rank-Deficient LOS Environment, Doctoral Thesis, Georgia Institute of Technology, August 2009.

[66] E. C. van der Meulen "Three-terminal communication channels," *Adv. Appl. Probab*, vol.3, pp. 120-154, 1971.

[67] Thomas M. Cover and Abbas A. El Gamal, "Capacity theorems for relay channel," *IEEE Trans.Inform.Theory*, vol. 25, no. 5, pp. 572-584, September 1979.

[68] P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Trans.Inform.Theory*, vol. 46, no. 2, pp. 388-404, March 2000.

[69] G. Kramer, M. Gastpar and P. Gupta, "Cooperative Strategies and Capacity Theorems for Relay Networks," *IEEE Trans.Inform.Theory*, vol. 51, no. 9, pp. 3037-3062, September 2005.

[70] N. Laneman and G. W. Wornell, "Energy-efficient antenna sharing and relaying for wireless networks," *Proc. IEEE Wireless Communications and Networkings Conf. (WCNC)*, Chicago, IL , September 2000.

[71] N. Laneman, D. N. C Tse and G. W. Wornell, "Distributed space-time coded protocols for exploiting cooperative diversity in wireless network," *IEEE Trans. Inform. Theory*, vol. 49, pp. 2415-2425, October 2003.

[72] N. Laneman, D. N. C Tse and G. W. Wornell, "Cooperative Diversity in Wireless Networks: Efficient Protocols and Outage Behaviour," *IEEE Trans.Inform.Theory*, vol. 50, pp. 3062-3080, December 2004.

[73] P. A Anghel, G. Leus, M. Kaveh, "Multi-user space time coding in cooperative networks," *Proc. IEEE ICASSP 2003*, vol. 4, pp. IV-73-IV-76, April 6-10 2003.

[74] S. Barbarossa, G. Scutari, "Distributed space-time coding strategies for wideband multihop networks: Regenerative vs non-regenerative relays," *Proc. IEEE ICASSP 2004*, Montreal , Canada , May 2004.

[75] S. Barbarossa, G. Scutari, "Distributed space-time coding strategies for multi hop networks," *Proc. IEEE ICC 2004*, Paris, France, June 2004.

[76] P. A. Anghel, M. Kaveh, "On the diversity of cooperative systems," *Proc. of IEEE ICASSP 2004*, vol. 4 , pp. 577-580, May 17-22 2004.

[77] T. E. Hunter and A. Nostrania "Diversity through coded cooperation," *IEEE Trans Wireless Commun*, submitted for publication.

[78] M. Janani, A. Hedayat, T. E. Hunter, A. Nostrania "Coded Cooperation in Wireless Communications: Space-Time Transmission and Iterative Decoding," *IEEE Trans Signal processing*, vol. 52, pp. 362-371, February 2004

[79] A. Stefanov and Elza Erkip "Cooperative Coding for Wireless Networks," *IEEE Trans. on Commun*, vol. 52, pp. 1470-1476, September 2004.

[80] J. Boyer , D. D. Falconer and H. Yanikomeroglu "Multihop Diversity in Wireless Relaying Channels," *IEEE Trans. on Commun*, vol. 52, pp. 1820-1830, October 2004.

[81] A. Sendonaris, E. Erkip and B. Aazhang, "User cooperation diversity, Part I: System description," *IEEE Trans. Inform. Theory*, vol.51, pp.1927-1938, November 2003.

[82] A. Sendonaris, E. Erkip and B. Aazhang, "User cooperation diversity, Part II: Implementational aspects and performance analysis," *IEEE Trans. Inform. Theory*, vol. 51, pp. 1939-1948, November 2003.

[83] I.Maric, R.D.Yates, "Forwarding strategies for Gaussian parallel relay networks," *Conference on Information Sciences and Systems CISS 2004*, 6 pages, 2004.

[84] B. Schein and R. Gallagher, "The Gaussian parallel relay network," *Proc. ISIT 2000*, Sorreno , Italy, 25-30 June 2000,

[85] R. Pabst, B.H. Walke, D.C. Schultz, P. Herhold, H. Yanikomeroglu, H. Viswanathan, S. Mukherjee, M. Lott, W. Zirwas, M. Dohler, H. Aghvami, D.D. Falconer, and G.P. Fettweis. "Relay-based deployment concepts for wireless and mobile broadband radio," IEEE Communications Magazine, vol. 42, no. 9, pp. 80.89, September 2004.

[86] "Performance Evaluation of a Fixed Relay Concept for Next Generation Wireless Systems," IEEE PIMRC 2004, pp. 744-751.

[87] H. Hu, H. Yanikomeroglu, David D. Falconer, and S. Periyalwar "Range Extension without Capacity Penalty in Cellular Networks with Digital Fixed Relays," IEEE GLOBECOM 2004, vol 5, pp. 3053-3057, 29 November - 3 December 2004.

[88] O. Oyman, "Opportunistic Scheduling and Spectrum Reuse in Relay-Based Cellular Networks," IEEE Transactions on Wireless Communications, vol 9, no. 3, pp. 1074-1085, March 2010.

[89] Haneda K, Kolmonen V, Riihonen T et al Evaluation of Relay Transmission in Outdoor-to-Indoor Propagation Channels. *European Cooperation in Science and Technology Workshop (COST) 2100* , Trondheim, Norway, 2008.

[90] T. Isatalo T, P. Lhdekorpi P, J. Lemplinen, 'Improving HSDPA Indoor Coverage and Throughput by Repeater and Dedicated Indoor System," *EURASIP Journal on Wireless Communications and Networking* Artcile ID 951481, 11 pages, 2008.

[91] A. Saleh, S. Redana, B. Raaf et.al "Performance of Amplify-and-Forward and Decode-and-Forward in LTE-Advanced ," *Proc. of IEEE VTC 2009* , pp. 1-5, Anchorage, 20-23 September 2009:

[92] H. Vishwanathan and S. Mukherjee "Performance of Cellular Networks with Relays and Centralized Scheduling," IEEE Transactions of Communications, vol. 4, Issue 5, pp. 2318-2328, November 2005.

[93] S. Peters, A. Y. Panah, K. T. Truong and R. W . Heath Jr. "Relay Architectures for 3GPP LTE-Advanced," *EURASIP Journal onWireless Communications and Networking*, vol. 2009, Article ID 618787, 14 pages, 2009. doi:10.1155/2009/618787.

[94] R. Irmer and F. Diehm "On Coverage and Capacity of Relaying in LTE-Advanced in Example Deployments," *Proc. of PIMRC 2008*, pp. 1-5, 15-18 September 2008.

[95] K. Bakanoglu, D. Gunduz and E. Erkip 'Dynamic Resource Allocation For The Broadband Relay Channel," *Proc. Asilomar 2008* , pp. 458-461, Pacific Grove, California, 4-7 November 2007.

[96] L. Huang, M. R. L Wang, Y. Xue and E. Schulz "Resource Allocation for OFDMA Based Relay Enhanced Cellular Networks," *Proc. IEEE VTC 2007-Spring* , pp.3160-3164, Dublin, April 2007.

[97] Y. Pan, A. Nix and M. Beach "Resource Allocation Techniques for OFDMA-Based Decode-and-Forward Relaying Networks," IEEE VTC 2008-Spring, pp. 1717 - 1721, Singapore, 1-14 May 2008.

[98] T. Girici "Fairness and QoS-Based Resource Allocation in Multihop Relay Networks," Proc. IEEE WCNC 2009.

[99] L. Xiao and L. Cuthbert "A Two-hop Proportional Fairness Scheduling Algorithm for Relay Based OFDMA Systems," Proc. of IEEE WiCom 2008, pp. 1-4, Dalian, 12-14 October 2008.

[100] G. Li and H. Liu "Resource Allocation for OFDMA Relay Networks With Fairness Constraints," IEEE Journal on Seleced Areas in Communications, vol. 24, no. 11, pp. 2061-2069, November 2006.

[101] S. Ren and K. Letaief "Optimal Effective Capacity for Cooperative Relay Networks With QoS Guarantees," IEEE ICC 2008-Spring, pp. 3725-3729, Beijing, 19-23 May 2008.

[102] C. Bae and D. H Cho "Fairness-Aware Adaptive Resource Allocation Scheme in Multihop OFDMA Systems," IEEE Communications Letters, vol. 11, no. 2, pp. 134-137, February 2007.

[103] M. Kaneko, P. Popovski, and K. Hayashi "Throughput-Guaranteed Resource-Allocation Algorithms for Relay-Aided Cellular OFDMA System," IEEE Transactions on Vehicular Technology, vol. 58, no. 4, pp. 1951 -1964.

[104] A. Volgenant "Solving the k-cardinatlity assignment problem by transformation," *European Journal of Operational Research* , pp. 322- 331, 2004.

[105] N. Karmakar, "A New Polynomial-Time Algorithm for Linear Programming," Annals of Computing Machinery, pp. 302-311, 1984.

[106] M. Meyer, H. Wiemann et. al. "ARQ Concept for the UMTS Long-Term Evolution," *Proceedings of IEEE VTC Spring-2006*, vol. 1, pp. 1-5, Melbourne, Australia, March 2006.

[107] R. Susitaival , M. Meyer "LTE Coverage improvement by TTI bundling," *Proceedings of IEEE VTC Spring-2009*, Barcelona, Spain, 2009.

[108] M. Simon and M. Alouini, *Digital Communications over Fading Channels: A Unified Approach to Performance Analysis* , New York:Wiley, 2002.

[109] S. Stiglmayr, M. Bossert, E. Costa "Adaptive Coding and Modulation in OFDM Systems using BICM and Rate-Compatible Punctured Codes," *Proceedings of European Wireless 2008*, Invited paper, Paris, 2007.

[110] F. Peng , J. Zhang, W. E. Ryan, "Adaptive Modulation and Coding for IEEE 802.11 n," *Proceedings of WCNC 2007*, pp. 656-661, Hong Kong, 2007.

[111] E. Akay, E. Ayanoglu "High Performance Viterbi Decoder for OFDM Systems," *Proceedings of IEEE VTC Spring-2004*, vol. 1, pp. 323-327, Milan, Italy, 2004.

[112] D.N.C Tse, P. Vishwanath. *Fundamentals of Wireless Communications.* Cambride University Press, 2005.

[113] S. Boyd, L. Vandenberghe, Convex optimisation, *Cambridge University Press*, 2004, New York.

[114] R. Schoenen, R. Halfmann, and B. H. Walke "An FDD Multihop Cellular Network for 3GPP-LTE," *Proc. IEEE VTC 2008*, pp. 1990-1994, Singapore, 11-14 May 2008.

[115] V. Venkatkumar and M. Faulkner "Energy-efficiency of Multiple-Relay Cooperation in Secsor Networks," RUNES Summer School, London, UK, 9-11 July 2007. http://www.ist-runes.org

[116] T. Haustein, V. Venkatkumar et. al, "Measurements of Multi-Antenna Gains using a 3GPP-LTE Air Interface in Typical Indoor and Outdoor Scenarios," *Proc. European Wireless 2008*, pp. 1-6, Prague, 22-25 June 2008.

[117] T. Wirth, V. Venkatkumar, T. Haustein, E. Schulz and R. Halfmann, "LTE-Advanced Relaying for Outdoor Range Extension," IEEE VTC2009-Fall, pp. 1-4, Anchorage, USA, 20-23 September 2009.

[118] V. Venkatkumar, T. Wirth, T. Haustein and E. Schulz, "Relaying in Long Term Evolution: Indoor full frequency reuse," IEEE European Wireless 2009, pp. 298-302, Aalborg, 17-20 May 2009.

[119] A. Forck, T. Haustein, V. Jungnickel, V. Venkatkumar, S. Wahls, T. Wirth, and E. Schulz, "Early real-time experiments and field-trial measurements with 3gpp-lte air interface implemented on reconfigurable hardware platform," 3GPP LTE Handbook: 3GPP LTE Radio and Cellular Technology, Editors B. Furht and S. A. Ahson ch. 11, pp. 365411, CRC Press, Auerbach Publications Boston, MA, USA, 2009.

[120] V. Venkatkumar, T. Haustein et. al "Field Trial Results on Multi-User MIMO Downlink OFDMA in Typical Outdoor Scenario Using Proportional Fair Scheduling," International ITG Workshop on Smart Antennas, 2008. WSA 2008, pp. 55-59, Darmstadt, 26-27 February 2008.

[121] T. Wirth, V. Venkatkumar et. al, "Polarisation Dependent MIMO Gains on Multiuser Downlink OFDMA with a 3GPP LTE Air Interface in Typical Urban Outdoor Scenarios," International ITG Workshop on Smart Antennas, 2008. WSA 2008, pp. 157-161, Darmstadt, 26-27 February 2008.