RESOURCE ALLOCATION ALGORITHMS FOR STATISTICAL QOS GUARANTEES IN MIMO CELLULAR NETWORKS

by

Mehmet Özerk Memiş

Submitted to the Graduate School of Engineering and Natural Sciences
in partial fulfillment of
the requirements for the degree of
Master of Science

Sabancı University

Summer 2013

RESOURCE ALLOCATION ALGORITHMS FOR STATISTICAL QOS GUARANTEES IN MIMO CELLULAR NETWORKS

APPROVED BY

Assoc. Prof. Dr. Özgür Erçetin ..... (Thesis Supervisor)

Assoc. Prof. Dr. Özgür Gürbüz ..... (Thesis Co-supervisor)

Assist. Prof. Dr. Hakan Erdoğan .....

Assoc. Prof. Dr. Barış Balcıoğlu .....

Assoc. Prof. Dr. Onur Kaya .....

DATE OF APPROVAL: .....

to my family

Acknowledgments

I would like to express my sincere gratitude to my thesis advisor, Assoc. Prof. Özgür Erçetin for his invaluable support, supervision and useful suggestions throughout this research. I am grateful to him for his infinite patience, valuable discussions and reviews which enabled me to complete my work successfully.

I would also like to thank my thesis co-advisor, Assoc. Prof. Dr. Özgür Gürbüz whose encouragement and guidance has helped me through all my research.

I would like to express my gratitudes to Assist. Prof. Dr. Hakan Erdoğan, Assoc. Prof. Dr. Barış Balcıoğlu and Assoc. Prof. Dr. Onur Kaya for their participation in my thesis committee and review of my thesis.

I would like to thank TÜBİTAK for providing me the financial support via BİDEB 2210 program.

Lastly, let me give my special thanks to my parents, sister and Ayşe for their concern, motivation and endless support. I am grateful to my family to let me go on my own way.

Abstract

Multiple-input-multiple-output (MIMO) antenna technology has attracted significant interest in recent years due to its great potential to increase wireless capacity and to provide reliability without extra power and/or bandwidth consumption. Thus, MIMO antenna technology finds wide employment in current wireless networking standards such as wireless LAN (IEEE 802.11n) and it is also expected to be employed in the next-generation systems such as 4G cellular networks. Moreover, as the diversity in services provided to mobile users increases, the capability to support diverse delay quality-of-service (QoS) requirements arises as a key feature of next-generation networks.

This thesis investigates resource allocation schemes in the downlink channel of MIMO cellular networks serving multiple users with different delay QoS requirements. This work specifically focuses on proportionally fair resource allocation algorithms that optimize the aggregate system utility given in terms of "effective capacity" of users. The effective capacity of a user identifies the maximum arrival rate supportable by the system while satisfying a probabilistic delay constraint. Resource allocation problem is solved for both time-division-multiple-access (TDMA) and space-division-multiple-access (SDMA) systems, and two resource allocation algorithms for each are given. In a TDMA system, each user is assigned a distinct slot of optimal length, based on the instantaneous channel conditions and QoS requirements of active users in each frame. In a SDMA system, multiple streams are transmitted simultaneously. The transmitter gives different power assignments to each stream determined as a solution to the utility maximization problem. The performance and the efficacy of the proposed algorithms are demonstrated both via numerical experiments and simulations considering realistic channel models and various QoS settings.

Özet

Çok Girişli Çok Çıkışlı (ÇGÇÇ) anten teknolojisi, ek güç ve/veya bant genişliğine ihtiyaç duyulmadan, kablosuz kanal kapasitesini artırması ve güvenilir iletişim sağlaması nedeniyle son yıllarda oldukça ilgi çekmektedir. Bu sebeple, ÇGÇÇ anten teknolojisinin kablosuz yerel ağlar (LAN) (802.11n) gibi günümüz kablosuz ağlarında geniş kullanım alanı bulduğu gibi, 4G gibi yeni-nesil hücresel ağlarda da kullanılması beklenmektedir. Son yıllarda, yeni-nesil ağların artan veri hızlarını desteklemelerinin yanısıra, türdeş olmayan rötar gibi kullanıcı servis kalitesi isterlerini karşılayabilmeleri önem kazanmaya ve kilit bir özellik olmaya başlamıştır.

Bu tezde, çok kullanıcılı hücresel ağlardaki uca yönelim sistemlerinde sınırlı zaman ve güç kaynaklarının eniyileme yöntemleriyle yönetimi, aynı anda ÇGÇÇ kanal getirilerinden faydalanılarak ve türdeş olmayan kullanıcı servis kalitesi isterleri göz önünde bulundurularak, kullanıcıların etkin kapasiteleri üzerinden tanımlı fayda işlevleri toplamları üzerinden orantılı adil servis sağlayan iki farklı temel algoritmayla değerlendirilmiştir; zaman paylaşımı algoritması (ZPA) ve güç paylaşımı algoritması (GPA). Kullanıcılara servis, ZPA'da dinamik olarak girişim sakınımlı ayrık zaman paylaşımlı ışın biçimlendirmeyle, GPA'da ise üst üste kodlama ile yapılan eş zamanlı iletimde güç kontrollü girişim yönetimiyle verilmektedir. ZPA'da her çerçevede tüm kullanıcıların kanal durumlarına bağlı olarak, her kullanıcıya en iyi uzunlukta ayrık zaman tahsisi yapılmaktadır. GPA'da ise kaynak paylaşımı tüm kullanıcıların kanal durumlarının uzun süreli ortalamaları göz önünde bulundurularak gerçekleştirilmektedir.

Çalışmada, gerçekçi kanal modelleri ile çeşitli kullanıcı servis kalitesi isterleri göz önünde bulundurularak alınan çözümsel verilerle ve yapılan benzetim ortamı deneyleriyle, önerilen algoritmaların etkinlikleri gösterilmiştir.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

Last-mile connections to end-users are becoming predominantly wireless. Rapid increase in use of mobile devices providing various real-time services such as IPTV, VoIP, Internet Radio, video conferencing results in search for better use of the wireless medium. In order to deliver the same performance to end-users as if they are connected to a wired network, new techniques to maximize the throughput in all-wireless networks must be developed. One of the most promising approaches in achieving this, is the use of multiple-input multiple-output, or MIMO, technology [1], [2]. In MIMO, both the transmitter and receiver are equipped with multiple antenna elements, where each antenna pair provides an independent spatial path between the transmitter and receiver. Hence, the capacity scales linearly with the number of antenna elements even though the antennas transmit and receive simultaneously on the same frequency band [3].

Besides increased data rates, the efficient use of the wireless medium also requires to take into account the heterogeneous quality-of-service (QoS) constraints (e.g. delay constraints) imposed by each different service provided by mobile network operators and internet service providers.

Independent of the wireless channel, to overcome latency in data transmission, data compression methods are used, which increase the information carried by unit time. In addition, multiple alternative paths, i.e. wireless channels, can be consid-

ered depending on the varying nature of the wireless medium, via multiple access points. In [4], these two approaches are combined to show that given a delay-QoS the efficient use of the wireless medium for video transmission can be improved. However, such approaches are limited in the sense that they cannot efficiently exploit the wireless channel characteristics. E.g. in terms of supporting high definition (HD) video streaming, there are works focused on transmission rate adaptation schemes, that take into consideration the variations in the wireless channel [5]. However, from the physical layer (PHY) perspective, these schemes are hard to employ in multi-user wireless networks, and thus, need further improvement.

In this thesis, we focus on downlink channel multi-user QoS provisioning via different resource allocation methods in MIMO cellular networks while considering both PHY and media access control (MAC) layers. There is a plethora of work on cross-layer resource optimization in wireless systems. All these works illustrate that significant throughput gain can be obtained by joint optimization of radio resource across PHY and MAC layers. A typical assumption is that the transmitter has an infinite backlog and the information flow is delay insensitive. However, in practice, it is very important to consider random bursty arrivals and delay performance metrics in addition to the conventional PHY layer performance metrics in cross-layer optimization.

To achieve efficient wireless communications while supporting diverse delay QoS requirements, we employ the effective capacity as the main performance metric in this thesis. The effective capacity was defined in [6] to evaluate the capability of a wireless service process in supporting data transmission subject to a statistical delay QoS requirement metric, called QoS exponent and denoted by θ. A higher θ corresponds to a more stringent delay constraint. Also, θ can continuously vary from 0 to ∞, and thus a wide spectrum of QoS constraints can be readily characterized by a general model. However, incorporating the effective capacity model into multi-user communications faces significant challenges, which are not encountered in a single user wireless link. Multi-user systems often have to "carefully" allocate the wireless resources based on mobile users' channel state information (CSI), and they usually need to balance the performances among all mobile users according to users' diverse

QoS requirements.

1.2 Literature Review

Existing practical wireless networks that are based on the multi-layer communication structure, provide modularity and transparency across the layers, which led to today's robust and flexible standard internet protocols [7]. However, the multiple layers functioning independent of each other, cause inefficient use of the wireless resources provided by MIMO systems. Exploiting the characteristics of MIMO technology in the physical layer and translating its performance gains to higher layers has motivated the integrated, cross layer approaches [8]. MIMO technology specifically requires designing and managing the interaction between the PHY and MAC layers. Additionally, fairness constraints together with QoS requirements imposed by the higher layers and time-sensitive applications must be taken into account in designing an integrated communication system.

The most fundamental unit used in resource management of MIMO systems is called a MIMO stream. A MIMO stream is logically defined as the spatial communication channel that is obtained by making use of one the spatial degrees-of-freedom (DoF) of the MIMO channel. Physically, MIMO streams are fundamental spatial channels that are obtained by cooperative coding of multi-antennas on both transmitter and receiver sides. Accurate computation of the MIMO channel capacity requires complex matrix operations and use of methods that do not provide closed-form solutions. This hardens the integration of the PHY models to the higher layers. In order to overcome this difficulty and enable cross-layers designs, closed-form, simpler and accurate uncorrelated MIMO channel capacity computation methods have been proposed [9]. There are also capacity computation methods that employ K-state Markov models and Gilbert-Elliot (GE) channel models for correlated MIMO channels [10]. In both approaches, MIMO channels are defined via their DoF, i.e., independent signaling dimensions.

The MIMO stream term has been employed in Stream Controlled Multiple Acces (SCMA) protocol, which provides resource management and adaptation of the higher

layers with respect to the MIMO channel in [11], [12]. The main principle of this system is to schedule links that can cause congestion without the stream approach, and then scheduling the remaining links based on streams. In the literature, optimal scheduling policies based on the stream-based structure and considering the trade off between spatial multiplexing and diversity have been proposed [13]. Another work focused on MIMO streams proposes a stream control approach that aims to increase the efficiency of a wireless network with a distributed and two-level heuristic [14].

[15] focuses on the admission control problem in transport and application layers of MIMO-based wireless networks. Reception and transmission capabilities of the nodes are quantified based on channel estimation errors, transmit power levels and channel statistics, and probabilistic DoF for spatial transmission and reception are extracted. The proposed model also considers the trade off between the spatial multiplexing and reuse advantages of MIMO, so that admission control allows higher flow rates with efficient resource utilization. However, the work focuses only on the admission control and it does not provide any scheduling and routing solutions.

As far as cross-layer MIMO resource management is concerned, time-division-multiple-access (TDMA) based interference aware transmission scheduling [16] is used to resolve contention problems in wireless networks by considering spatial DoF [17]. There are also works, such as [18] that take into account scheduling, transmission power control and routing in an integrated manner. This work also incorporates data rate and queue stability constraints, resulting in a large and complex optimization problem, which can only be solved by dividing the network into broadcast domains and applying dynamic programming methods within each broadcast domain separately. However, the performance enhancements of the proposed method are not quantified in a realistic simulation environment.

Opportunistic scheduling policies are also proposed for the channel optimization of MIMO based wireless networks [19]. The works in the literature show that the use of opportunistic MAC protocols together with multi-user MIMO systems can increase the system capacity up to threefold in wireless mesh networks [20]. A similar opportunistic approach is proposed together with cooperative methods [21]. The work reports that the proposed TDMA-based special scheduling method can

achieve eight times better network capacity values than standard 802.11 networks.

In 4G wireless networks, maximizing the system capacity while satisfying QoS requirements of different user applications is crucial. In addition to works that consider different fairness criteria in scheduling, such as [22], there are a limited number of studies on scheduling in MIMO networks with QoS constraints. The main approach common on all these works is that packet transmission schedules and resource allocation are handled in the MAC layer, and PHY layer performs beamforming and multiuser diversity gain. [23] proposes such an approach, with a packet prioritizer followed by a resource allocator, which tries to maximize the throughput of a given packet priority order.

There are various approaches dealing with delay-QoS-aware resource allocation control in wireless networks [24], [25]. One approach dealing with delay-QoS-aware resource control employs the notion of Lyapunov stability and in the stability sense builds throughput optimal control policies. The throughput optimal policies guarantee the stability of the queueing network if stability is to be achieved under any policy. Three known throughput-optimal classes of policies are the Max Weight rule [26], the Exponential rule [27], and the Log rule [28]. Especially, the Max Weight-type class of algorithms, which are proved to minimize the Lyapunov drift and are throughput-optimal, are utilized in many dynamic control algorithms used to optimally allocate limited resources in satellite and wireless systems [29], [30], [31].

A second and more systematic approach to deal with delay-QoS-optimal resource allocation control is the Markov decision process (MDP) approach. It is possible to obtain delay-optimal solutions in some special cases, e.g. [32], [33], where the authors show that the longest queue highest possible rate (LQHPR) policy is delay-optimal for multi-access systems with homogeneous users. However, in general this approach is not utilized due to the challenges mentioned in [34].

A third and most-widely used approach in dealing with delay-aware resource control is to utilize effective bandwidth and effective capacity theories, which convert average delay constraints into equivalent average rate constraints using large deviation theory. In this approach, the optimal resource allocation problem is solved using an information theoretical formulation based on the average rate con-

straints [35], [36], [37], [38], [39].

The key aspect of guaranteeing delay-QoS in a wireless network in the third approach is to be able to model both the data arrival traffic and the service offered by the network, i.e. wireless channel process. The majority of the works utilizing this approach like [40], [41], where trade-off between power allocation and delay is studied, consider single-input-single-output (SISO) wireless channels due to relative simplicity of extracting effective capacity expressions. However, when MIMO channels are considered, even the modeling of a single wireless channel process turns into a challenging problem [42]. Due to these challenges, closed form solutions for optimal resource allocation under QoS provisioning in multi-user MIMO networks rarely exist [43], which are usually hard to be utilized in real systems.

A cross-layer design controlling MAC and PHY layers is essential for designing optimal scheduling for MIMO systems. Exploiting the physical characteristics and flexibility of MIMO while satisfying individual users' QoS requirements remains to be an active research topic.

1.3 Problem Statement

In this thesis, we consider a single cell of a MIMO cellular network, where the base station in each cell operates in a different frequency band than the base stations in the neighboring cells within its transmission range. We investigate two different resource allocation schemes for MIMO users receiving delay-sensitive data streams from the base station over time-varying wireless channels.

In the first approach, we adopt an interference-free model, where only one user transmits at a time. The base station acquires the instantaneous channel state information from each user and determines how long each user receives service from the base station within a time frame. We refer to this resource allocation scheme as resource allocation in time-division-multiple-access (TDMA) system.

In the second approach, all users are served simultaneously by precoding the data streams prior to transmission. In this case, transmission power of each of the user streams are determined based on the average channel state information.

This scheme is referred to as resource allocation in space-division-multiple-access (SDMA) system.

These two approaches complement each other, since the acquisition of instantaneous CSI may induce large overheads, and thus, it may be prohibitively expensive to be implemented in some systems. On the other hand, precoding of data streams require more complex wireless transceivers, which may not be preferred due to cost considerations.

In both of these approaches, we first model the effective capacity of MIMO links by explicitly taking into account multi-user scheduling and resource allocation. Based on this model, we formulate the resource allocation problem as a network utility maximization (NUM) problem with each user having potentially different quality of service requirement. The solution of this problem under realistic channel models and the efficacy of the algorithms are demonstrated first by numerical experiments, and then via network level simulations.

1.4 Contributions and Thesis Organization

Our main contributions in this thesis can be summarized as follows:

- We propose two different proportionally fair resource allocation algorithms in form of NUM problems for both the TDMA system and the SDMA system. The objective in the proposed algorithms is to optimize the aggregate system utility given in terms of effective capacity of users in the downlink channel of MIMO cellular networks. The base station within each cell has limited wireless resources and serves multiple users with different delay-QoS requirements .

- We show the performance and the efficacy of the resource allocation schemes via both numerical analysis and simulations under realistic channel models and considering various delay-QoS requirements compared to trivial equal resource allocation schemes.

- We propose a simple but accurate closed-form moment generating function (MGF) and effective capacity expression for MIMO channel process, which is

obtained by discretization of the MIMO channel process.

- We propose an approximate effective capacity expression as a function of power allocation vector by forming distribution models for the eigenvalue distributions of the MIMO channel matrices and employing central limit theorem (CLT), which is utilized in the practical power allocation algorithm that is proposed for resource allocation in the SDMA system.

The organization of the thesis is as follows. In Chapter 2, a general background is given on the theories and technologies utilized in the proposed algorithms. In Chapter 3, we first introduce our closed-form effective capacity expression obtained by state-aggregation of the MIMO channel process, present static and dynamic frame allocation algorithms, and show their performance via numerical analysis. Chapter 4 introduces a static power allocation algorithm and a practical algorithm that uses the approximate effective capacity expression for the MIMO channel process as a function of power allocation vector, and concludes with their numerical analysis. Chapter 5 explains briefly the simulation environment and presents simulation data obtained with ns-2 to show the performance of the algorithms in a realistic setting and provides delay-QoS analysis for users with different QoS demands under various channel conditions. The thesis ends with Chapter 6, which contains the conclusion and the planned future work.

Chapter 2

Background

2.1 Effective Bandwidth and Effective Capacity Theory

As a need to analyze the delay control problem in asynchronous transfer mode (ATM) and internet protocol (IP) networks, the authors formed [44], [45] and extensively used [46], [47] the effective bandwidth theory, which models the asymptotic stochastic behavior of source traffic process to a queueing system, and tries to figure out the minimum constant channel rate that can serve a stationary source process while guaranteeing a target delay-QoS requirement, such that the delay does not exceed a given bound D_t with probability, $(1-\epsilon)$.

Consider a single queue system with instantaneous arrival rate $a(\tau)$ and channel service $c(\tau)$ in terms of bits, which arrive at and served by the queue in a finite length slot of τ seconds, respectively. Let $A(t)$ be the cumulative source process, i.e. the aggregate number of bits that arrived at the queue in $[0,t]$ expressed as $A(t) = \sum_{\tau=0}^{t} a(\tau)$, and $C(t)$ denote the cumulative channel process, i.e. aggregate number of bits served by the queue in $[0,t]$ expressed as $C(t) = \sum_{\tau=0}^{t} c(\tau)$. Define the workload process for the queue as $Q(t) = (A(t) - C(t))^+$ with $(x)^+ \triangleq max(0,x)$ and provided that $c(t) \leq Q(t) \quad \forall t$, which means that at any instant there are bits

to be transmitted. Next, we define the Gartner-Ellis limit of $Q(t)$ by

$$\alpha_Q(\theta) = \lim_{t\to\infty} \frac{1}{t} \log \mathbb{E}\left[e^{\theta Q(t)}\right]. \tag{2.1}$$

Exploiting the independence of both $a(\tau)$ and $c(\tau)$, $\alpha_Q(\theta)$ can be decomposed into two terms, i.e.

$$\alpha_Q(\theta) = \alpha_A(\theta) + \alpha_C(-\theta). \tag{2.2}$$

Defining $E_Q(\theta) = \alpha(\theta)/\theta$ (i.e. effective bandwidth function definition in [6]) we get

$$E_Q(\theta) = E_B(\theta) - E_C(\theta). \tag{2.3}$$

Now, we focus on the terms $E_B(\theta)$ and $E_C(\theta)$.

Assuming that the Gartner-Ellis limit of $A(t)$, denoted by

$$\alpha_B(\theta) = \lim_{t\to\infty} \frac{1}{t} \log \mathbb{E}\left[e^{\theta A(t)}\right], \tag{2.4}$$

exists for all $\theta \geq 0$, *the effective bandwidth function* of $A(t)$ is defined as

$$E_B(\theta) = \frac{\alpha_B(\theta)}{\theta} = \lim_{t\to\infty} \frac{1}{\theta t} \log \mathbb{E}\left[e^{\theta A(t)}\right]. \tag{2.5}$$

Now, considering a queue with infinite buffer size served by a channel with a constant service rate R, it is shown [44] that the probability of the instant delay $D(t)$ exceeding a delay bound D_t, i.e. target delay, satisfies

$$\epsilon = \sup_t \mathbb{P}\{D(t) > D_t\} = \gamma(R).e^{-\theta(R).D_t}. \tag{2.6}$$

by the large deviation theory [48], where $\gamma(R) = \mathbb{P}\{D(t) \geq 0\}$ is the probability that the queue is not empty and $\theta(R) = R.E_B^{-1}(R)$ is the so-called QoS exponent, i.e. R multiplied by the solution of $E_B(\theta) = R$. $\theta(R)$ is used as the metric for QoS requirement, such that a higher $\theta(R)$ indicates a stricter QoS requirement and vice versa. Since both $\theta(R)$ and $\gamma(R)$ are functions of constant channel rate R, a

source with a common delay bound D_t is said to be able to tolerate a delay violation probability of at most ϵ with the channel capacity being at least R. This means that the tail probability $\mathbb{P}\{D(t) > D_t\}$ is proportional to the queue being nonempty and decays exponentially fast as D_t increases.

Inspired by the effective bandwidth theory, where a constant channel rate is used to model the source traffic in wired networks, the authors in [6] used a constant source traffic rate μ and developed a dual effective capacity theory, in order to analyze the random and time-varying wireless communication channel. Contrary to effective bandwidth theory, the effective capacity theory tries to figure out the maximum constant arrival rate that can be served by a stationary channel service process at a queue, while satisfying a target delay-QoS requirement, such that the delay does not exceed a given bound D_t with probability, $(1 - \epsilon)$.

Assuming that the Gartner-Ellis limit of $C(t)$, denoted by

$$\alpha_C(-\theta) = \lim_{t\to\infty} \frac{1}{t} \log \mathbb{E}\left[e^{-\theta C(t)}\right], \tag{2.7}$$

exists for all $\theta \geq 0$, the effective bandwidth function of $C(t)$, i.e. *effective capacity function*, is defined as

$$E_C(\theta) = -\frac{\alpha_C(-\theta)}{\theta} = -\lim_{t\to\infty} \frac{1}{\theta t} \log \mathbb{E}\left[e^{-\theta C(t)}\right]. \tag{2.8}$$

Note that if the process $C(t)$ is uncorrelated, the effective capacity reduces to

$$E_C(\theta) = -\frac{1}{\theta} \log \mathbb{E}\left[e^{-\theta c(t)}\right]. \tag{2.9}$$

Now, considering again queue with infinite buffer size served by a data source with a constant data rate μ rate, it is shown [6] that the probability of the instant delay $D(t)$ exceeding a delay bound D_t, i.e. target delay, satisfies

$$\epsilon = \sup_t \mathbb{P}\{D(t) > D_t\} = \gamma(\mu).e^{-\theta(\mu).D_t} \tag{2.10}$$

by the large deviation theory [48], where $\gamma(\mu) = \mathbb{P}\{D(t) \geq 0\}$ is the probability that

the queue is not empty and $\theta(\mu) = \mu.E_C^{-1}(\mu)$ is the so-called QoS exponent. Since both $\theta(\mu)$ and $\gamma(\mu)$ are functions of constant source rate μ, a source with a common delay bound D_t is said to be able to tolerate a delay violation probability of at most ϵ with the data rate being at most μ. As in effective bandwidth theory, a high QoS exponent indicates a strict QoS requirement.

2.2 MIMO Channels

Since its introduction [1], [2], MIMO technology is widely employed as a key feature to increase wireless capacity. By using multiple antennas on both the transmitter and the receiver side of a wireless channel, it is shown that the wireless capacity can increase almost linearly [3]. Due to this critical gain introduced by MIMO technology into the wireless communications, it is widely utilized in wireless LAN (IEEE 802.11n), WiMAX access networks (IEEE 802.16), and 4G cellular networks (LTE). This performance gain, called as multiplexing gain, is a result of the fact that a MIMO channel can be decomposed into a number of independent parallel channels, the number of which depends on the number of antennas employed on both sides of a link. Multiplexing independent data onto these independent channels results in a linear increase in data rate compared to a system with one antenna at both transmitter and receiver side.

The channel of a MIMO link l with n_t transmit and n_r receive antennas is characterized by $\boldsymbol{H}^l$, i.e. the $n_r \times n_t$ channel gain matrix. The channel gain matrix consists of elements h_{ij}^l with i and j denoting the row and column indices, respectively. Such a channel is said to have a maximum degrees of freedom (DoF) of $d = \min\{n_t, n_r\}$, i.e. the maximum number of independent signaling dimensions. Communication over a wireless MIMO channel with $\boldsymbol{H}^l$ is described by

$$\boldsymbol{y}^l = \boldsymbol{H}^l \boldsymbol{x}^l + \boldsymbol{n}^l, \tag{2.11}$$

where $\boldsymbol{x}^l$, $\boldsymbol{y}^l$ and $\boldsymbol{n}^l$, represent the vectors of transmitted signal, received signal and zero-mean white Gaussian noise with variance σ_n^2. The entries of the channel matrix

$\boldsymbol{H}^l$, i.e. h_{ij}^l, denote the channel gain between the i^{th} antenna of the transmitter and the j^{th} antenna of the receiver.

To compute the MIMO channel capacity, the diagonalization of the channel gain matrix $\boldsymbol{H}^l$ is required, in which the channel is transformed into a set of parallel spatial channels. By singular value decomposition (SVD) of the channel gain matrix, $\boldsymbol{H}^l$ is written as

$$\boldsymbol{H}^l = \boldsymbol{U}^l \boldsymbol{\Sigma}^l \hat{\boldsymbol{V}}^l, \tag{2.12}$$

where $\boldsymbol{U}^l$ and $\boldsymbol{V}^l$ are unitary matrices, $\hat{x}$ denotes the Hermitian transpose, and $\boldsymbol{\Sigma}^l$ is a diagonal matrix with the singular values σ_i^l of $\boldsymbol{H}^l$ on its diagonal. If $\boldsymbol{H}^l$ is a random matrix, $\boldsymbol{\Sigma}^l$ may change its size depending on the number of non-zero singular values at a given instance. The number of non-zero singular values of $\boldsymbol{H}^l$ is referred to as the available DoF of the MIMO link.

The parallel decomposition of the MIMO channel is obtained by defining a linear transformation on the channel input vector $\boldsymbol{x}^l$ as $\tilde{\boldsymbol{x}}^l = \hat{\boldsymbol{V}}^l \boldsymbol{x}^l$, i.e. transmit precoding, and another linear transformation on the channel output vector $\boldsymbol{y}^l$ as $\tilde{\boldsymbol{y}}^l = \hat{\boldsymbol{U}}^l \boldsymbol{y}^l$, i.e. receiver shaping. Performing the mentioned linear transformation on the output vector, we get

$$\begin{aligned} \tilde{\boldsymbol{y}}^l &= \hat{\boldsymbol{U}}^l (\boldsymbol{H}^l \boldsymbol{x}^l + \boldsymbol{n}^l) \\ &= \hat{\boldsymbol{U}}^l (\boldsymbol{U}^l \boldsymbol{\Sigma}^l \hat{\boldsymbol{V}}^l \boldsymbol{x}^l + \boldsymbol{n}^l) \\ &= \boldsymbol{\Sigma}^l \tilde{\boldsymbol{x}}^l + \tilde{\boldsymbol{n}}^l. \end{aligned} \tag{2.13}$$

After decomposition, communication over the independent parallel channels is described by

$$\tilde{y}_i^l = \sigma_i^l \tilde{x}_i^l + \tilde{n}_i^l \quad \text{for} \quad i = 1, \ldots, d\,. \tag{2.14}$$

For a static MIMO channel, its capacity as the sum of the capacities of the

parallel channels is given by

$$C^l = \sum_{i=1}^{d} \log_2 \left[1 + \frac{P_i {\sigma_i^l}^2}{\sigma_n^2} \right], \tag{2.15}$$

when channel state is known at the transmitter side of the link. Capacity maximizing P_i values are found by water-filling (WF) algorithm by

$$P_i = \left(\mu - \frac{\sigma_n^2}{{\sigma_i^l}^2} \right)^+ \quad \text{for} \quad i = 1, \ldots, d\,. \tag{2.16}$$

Above, $(x)^+$ represents $max(0, x)$, μ denotes the water level computed by WF and P_i satisfy $\sum_i P_i \leq P$, where P is the total transmit power of the link.

In fading MIMO channels, the channel matrix entries h_{ij}^l vary with time. As in the case of the static channel, the instantaneous capacity of the link depends on what information on the channel matrix $\boldsymbol{H}^l$ is available at the transmitter side. Based on the available information, the transmitter can adapt to channel fading. In this case, the MIMO channel capacity is computed as the average of all channel matrix realizations with optimal power allocation, which is termed by ergodic capacity [49]. The ergodic capacity of the MIMO channel is given by

$$C_{erg}^l = \mathbb{E}_{\boldsymbol{H}^l} \left[\max_{P_i : \sum_i P_i \leq P} \sum_{i=1}^{d} \log_2 \left[1 + \frac{P_i {\sigma_i^l}^2}{\sigma_n^2} \right] \right]. \tag{2.17}$$

The capacity unit in the capacity expressions for both the static and fading MIMO channels is $bits/second/Hz$.

2.3 Superposition Coding

In the context of MIMO fading channels, superposition coding together with rate and power allocation has been applied to maximize the average transmission rate [50]. In superposition coding the encoder constructs the signals in a nested fashion in which the codeword, i.e. signal, that is intended for a certain receiver is a "satellite" of the codeword that is intended for the next more degraded receiver.

Let us first consider the two receiver case, and a scenario, where the signal observed by receiver 2 is more degraded than that observed by receiver 1. The transmitter wishes to communicate two independent messages simultaneously to both receivers. To do so, the transmitter synthesizes the signal, $\boldsymbol{x}$, by superimposing the signal $\boldsymbol{v}$, which contains the message intended for receiver 1 on the signal $\boldsymbol{u}$, which contains the message intended for receiver 2. The signal $\boldsymbol{u}$ is typically visualized as the center of a cluster of codewords and is chosen from a codebook with rate R^2. In each cluster, there are $(2)^{nR^1}$ satellites centered around $\boldsymbol{u}$, where n is the length of the codeword and R^1 is the rate of the codebook used for receiver 1. For Gaussian channels, when the transmit power budget is P, it was shown that the capacity achieving codebooks are independent and Gaussian, and that the average powers with which these codebooks are transmitted are $(1-\beta).P$ and $\beta.P$, where $\beta \in [0,1]$ is a partition of power among codebooks.

The decoding of superposition encoded signals, i.e. successive interference cancellation [51], is as follows. The Gaussian signal $\boldsymbol{v}$ contains the message intended for receiver 1. When operating at the boundary of the capacity region, this signal is not decodable by receiver 2, and hence receiver 2 sees it as additive Gaussian noise. Thus from receiver 2's perspective, the situation resembles an additive white Gaussian noise (AWGN) channel with signal power $\beta||\boldsymbol{H}^2||^2P$ and noise variance $\sigma_n^2+(1-\beta)||\boldsymbol{H}^2||^2P$. For receiver 2 to decode the signal $\boldsymbol{u}$, the rate R^2 must satisfy

$$R^2 \leq \log\left[1+\frac{\beta P||\boldsymbol{H}^2||^2}{(1-\beta)P||\boldsymbol{H}^2||^2+\sigma_n^2}\right]. \tag{2.18}$$

Since receiver 1 observes a channel that is less degraded than the channel observed by receiver 2, it can decode the signal $\boldsymbol{u}$, and subtract it from its received signal. Having done that, receiver 1 has a signal of power $(1-\beta)||\boldsymbol{H}^1||^2P$, and noise variance σ_n^2. Similarly, receiver 1 can correctly decode signal $\boldsymbol{v}$, if

$$R^1 \leq \log\left[1+\frac{(1-\beta)P||\boldsymbol{H}^1||^2}{\sigma_n^2}\right]. \tag{2.19}$$

For the base station to send independent messages to $L>2$ receivers, it generates L independent Gaussian codebooks, one for each degradation level [52]. The

transmitter superimposes L codewords, one from each codebook, to generate the transmitted signal. The transmitted signal can be regarded as a codeword from nested clusters. Each codebook represents a set of cluster centers that are decodable by the receiver at the corresponding degradation level as well as less degraded receivers. For more degraded receivers, these cluster centers are observed as undecodable satellites that contribute to the total noise observed by these receivers. Let ψ^l denote the particular degradation level of receiver l. The receivers at degradation levels $\psi^k < \psi^l$ are considered as less degraded receivers.

As codewords are transmitted from the nested clusters, the transmitter partitions its power, and in order to decode superposition-coded messages, each receiver begins by decoding and subtracting the signals intended for more degraded receivers. Treating the signals intended for less degraded receivers as additive Gaussian noise, each receiver then proceeds to decode its intended signal. Given a power partition $\boldsymbol{\beta} = (\beta^1, \ldots, \beta^L)$, and degradation levels ψ^l, for all $l = 1, \ldots, L$, the lth receiver is able to decode its intended signal, if the rate of the corresponding codebook satisfies:

$$R^l(\boldsymbol{\beta}) \leq \log \left[1 + \frac{\beta^l P ||\boldsymbol{H}^l||^2}{\sum_{j=1}^{L} \mathcal{I}_{\psi^j < \psi^l} \beta^j P ||\boldsymbol{H}^l||^2 + \sigma_n^2} \right], \tag{2.20}$$

where β^l is the partition of power allocated for user l, and $\mathcal{I}_{x<y}$ is an indicator function which takes value 1 when $x < y$, and 0 otherwise.

In this two receiver case, the signal observed by receiver 2, which is more degraded than the signal observed by user 1, is said to be encoded in the second position, and the signal observed by receiver 1 is said to be encoded firstly by the transmitter.

Chapter 3

Resource Allocation in TDMA System

In resource allocation in time-division-multiple-access (TDMA) system, we investigate exploitation of beamforming capability of MIMO communication, as the base station communicates with one user at a time over a point-to-point MIMO link in time division multiple access mode. Interference between users is avoided, since a particular portion, Φ^l, of the unit time frame is allocated to each user, i.e., each MIMO link, l, so that $\sum_{l=1}^{L} \Phi^l = 1$. Time durations to be allocated for the active users are variable size slots, which are changed dynamically, frame-by-frame, considering the users' QoS constraints and instantaneous channel conditions that is reflected to the available DoFs of the links.

For this problem, we define two algorithms, i.e. static resource allocation algorithm and dynamic resource allocation algorithm. In static resource allocation, we solve a one-shot NUM problem, with user utilities given as functions of effective capacities of the corresponding MIMO links. To solve this problem, we first derive an expression for the effective capacity of a single MIMO link. In dynamic resource allocation, we iteratively solve a NUM problem, where user utilities are denoted by functions of auxiliary variables, which are obtained from the derived effective capacity expression and are directly dependent on the instantaneously available DoF of the links.

3.1 System Model

We consider the downlink channel of a single cell in a MIMO based cellular network, where the base station is deployed with n_t antennas to communicate with multiple receivers each with n_r antennas. We assume a Gaussian broadcast scenario, in which the base station is sending independent messages to L receivers in time division multiple access mode with transmit beamforming, and the channel gain matrix observed by each receiver l is denoted by $\boldsymbol{H}^l$, which consists of circularly symmetric complex Gaussian (CSCG) entries $\boldsymbol{H}^l(i,j) = h_{ij}^l \sim \mathcal{CN}(0, {\sigma_{ij}^l}^2)$. The medium is assumed to have zero-mean Gaussian noise with variance σ_n^2, and the base station with both full and instantaneous channel state information (CSI) has a total transmit power of P Watts.

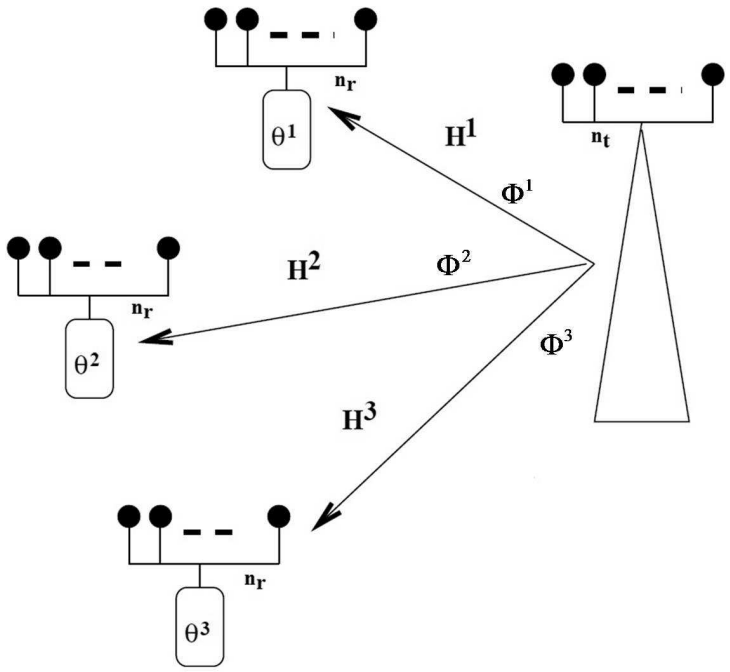


Figure 3.1: Resource allocation in TDMA system with three users.

As explained in detail in the following sections of this chapter, the proposed algorithms are designed with the aim of maximizing the aggregate system utility given in terms of effective capacity of L active users in the downlink channel. Each user has a different QoS exponent, θ^l, $l = 1, \ldots, L$.

In static resource allocation, the users obtain a utility which is a concave function of their effective capacity. Similarly, the utility that the users get in dynamic resource

allocation is a concave function of the derived auxiliary function that is a part of the effective capacity expression. In this work, we assume a logarithmic utility function which is shown to achieve proportional fairness among the users [53].

3.1.1 Channel Model

In order to to find the effective capacity of a MIMO channel process (i.e. (2.8)), one needs to both find a closed-form expression for the instantaneous MIMO channel rate and fully characterize its distribution, which involves an expectation operation with complex matrix operations to be performed together with water-filling algorithm. To overcome this complexity, the MIMO channel rate is expressed over its available Degrees of Freedom (DoF) [54].

This approach enables the point-to-point MIMO channel to be modeled as a discrete Markov chain, where each state i represents the number of available DoF with $i = 1, \ldots, d$ and $d = \min\{n_t, n_r\}$, and occurs with probability π_i [10]. For each link l, the total average Signal to Noise Ratio (SNR) is $\bar{\rho}^l = P.{\sigma_{ij}^l}^2/\sigma_n^2$. Given the average SNR of each link, $\bar{\rho}^l$, the discretized channel model can be obtained by considering sufficiently large number of channel realizations, applying singular value decomposition and water filling algorithm [1] for each channel matrix $\boldsymbol{H}^l$, marking the number of values exceeding the water level as the available DoF of the link, and then counting the occurrences of the different DoF to obtain the probability of lth link having i DoF, i.e., π_i^l, for all $i = 1, \ldots, d$.

Figure 3.2 displays π_i^l for a MIMO channel with $d = 3$ as a function of ${\sigma_{ij}^l}^2$ with $P/\sigma_n^2 = 5$, which are obtained utilizing the above mentioned method. As the channel gain increases, the probability of the MIMO channel having 3 DoF at a given frame increases and the probability of the MIMO channel having 1 DoF decreases, which is an expected result.

In resource allocation in TDMA system, active users are served once in each frame, which is of unit length normalized with respect to the channel coherence time. Hence, the available DoF and the total average SNR $\bar{\rho}^l$ per link remain constant throughout a frame. Due to fading, however, the available DoF per link can change independently from one frame to another, since there is no correlation

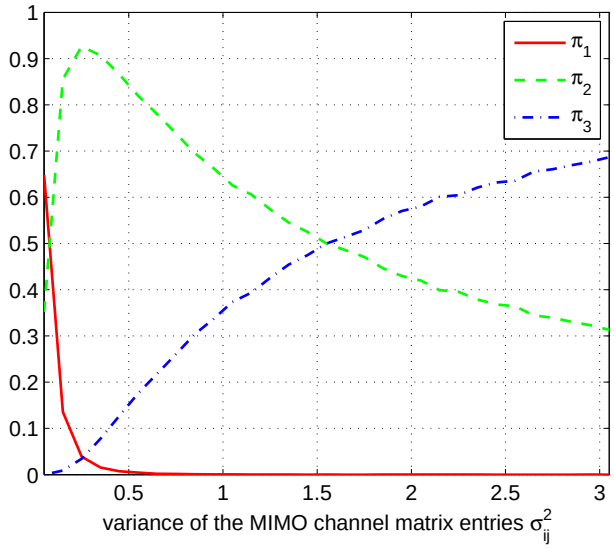


Figure 3.2: DoF probabilities, π_i for $i = 1, 2, 3$

between the entries of the channel gain matrices between two consecutive frames and the links are memoryless. With this assumption, (2.9) can be used.

The full CSI assumption for the base station means that it knows the Markov characterization of each MIMO link, and the instantaneous CSI assumption means that it knows the available DoF for each link in a given time frame.

Having introduced the discretization of the MIMO channel process, we now state the capacity of a MIMO link l in state i, that is approximately given as

$$R_i^l = i \cdot \log_2 \left[1 + \frac{\bar{\rho}^l}{i} \right] \tag{3.1}$$

with units bps/Hz. This is a simple, but accurate estimation of the ergodic (optimal) MIMO channel capacity obtained after singular value decomposition and water filling [54].

3.1.2 Effective Capacity Formulation

In order to calculate the effective capacity of a single MIMO link, we first determine the moment generating function (MGF) of the channel process. Note that the cumulative channel process of each MIMO link can be described as an uncorrelated homogeneous Markov Modulated Process (MMP). For a general MMP, MGF is

given by $\boldsymbol{\pi}(\Gamma(\theta)\mathbf{Q})^{t-1}\Gamma(\theta)\mathbf{1}^T$, where $\boldsymbol{\pi}$ is the steady-state probability vector, $\Gamma(\theta) = diag(e^{\theta R^0}, \ldots, e^{\theta R^d})$ is the rate matrix, $\mathbf{Q}$ is the state transition matrix and $\mathbf{1}$ is the column vector of ones [55]. Utilizing this expression and applying its definition, the MGF of the point-to-point MIMO channel process is determined as

$$M_C(\theta^l, t) = \mathbb{E}\left[e^{-\theta^l C(t)}\right] = \sum_{i=0}^{d} (\pi_i^l)^t e^{-\theta^l R_i^l t}\,, \tag{3.2}$$

where R_i^l is the transmission rate of MIMO link l when it has i DoF. Note that (3.2) reduces to the MGF of the ON-OFF traffic source, when a MIMO link has one antenna at both transmitter and receiver side [55].

Once the MGF of the service process is determined, the Gartner-Ellis limit of *log-MGF* can be calculated according to (2.7). However, due to the complexity of obtaining a closed-form expression, we use the approach presented in [56], which first finds an upper-bound on the *MGF* of the channel process and then extracts the effective capacity expression for a general channel process using (2.8), so that the given QoS constraint in form of QoS-exponent is not violated [57], i.e.

$$\begin{aligned}\log(M_C(\theta^l, t)) &= \log\left(\sum_{i=0}^{d} e^{t(\log \pi_i^l - \theta R_i^l)}\right) \\ &\leq \log\left((d+1)e^{t max_i\{\log \pi_i^l - \theta R_i^l\}}\right).\end{aligned} \tag{3.3}$$

Then, substituting (3.3) into (2.7), we obtain,

$$\alpha_C(-\theta^l) = \max_i\{\log \pi_i^l - \theta R_i^l\}. \tag{3.4}$$

Finally, the effective capacity of a single MIMO link with i DoF is obtained by substituting (3.4) in (2.9):

$$E_C^l(\theta^l) = \min_{i=0,\ldots,d}\left\{R_i^l - \frac{\log \pi_i^l}{\theta^l}\right\}. \tag{3.5}$$

The above given expression is valid for θ^l such that $E_C^l(\theta^l)$ does not exceed the Shannon capacity limit of the link.

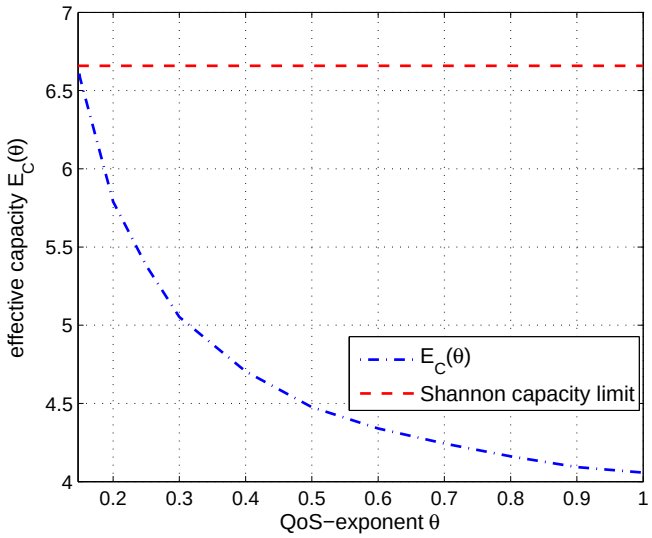


Figure 3.3: Effective capacity and Shannon capacity limit

In Figure 3.3, the effective capacity of a MIMO channel with $d = 3$ and $\bar{\rho}^l = 5$ is given. It is seen that for $\theta^l = 0.147$, the effective capacity is at the Shannon capacity limit 6.6583 bps/Hz for the given conditions. As the QoS requirement gets stricter, the effective capacity decreases, which means that in order for the wireless channel to satisfy a stricter QoS demand, it should operate at a lower rate.

3.2 Static Resource Allocation Algorithm

In static resource allocation in TDMA system, for each MIMO link a fraction of time is reserved at each time frame depending on the instantaneous DoF of all links and considering their QoS requirements in the system. The problem of resource allocation is solved once by taking into consideration all possible DoF combination of the links. With the solution of the NUM problem, what fraction of the frame is to be allocated to which user is known given all possible DoF combinations.

Let $\delta^l(t)$ be the available DoF in frame t and $\phi^l(t)$ represent the fraction of frame t reserved for this MIMO link. Then, the instantaneous transmission rate of a MIMO link l can be approximately given as:

$$R^l(t) = \delta^l(t)\,\phi^l(t)\,\log_2\left[1 + \frac{\bar{\rho}^l}{\delta^l(t)}\right]. \tag{3.6}$$

In order to determine the effective capacity of each link according to (3.5), we require the long term average rates for each DoF, $\mathbb{E}\left[R^l(t)|\delta^l(t)=i\right]$, $\forall i=1,\ldots,d$. The expectation is with respect to the joint channel state distribution of all links, which is simply the product of marginal distributions of individual MIMO links, since all links are independent. Hence, the problem reduces to the allocation of proportion of the time frame based on the DoF of each of the MIMO links given the joint channel distribution and QoS parameters θ^l. Let $\boldsymbol{\delta}=(\delta^1,\ldots,\delta^L)$ be the vector of DoFs of MIMO links, and $\Phi^l(\boldsymbol{\delta})$ be the proportion of frame allocated to link l when MIMO links have DoFs $\boldsymbol{\delta}$. Note that the effective capacity $E_C^l(\theta^l)$ is hence given by

$$\min_{i=1,\ldots,d}\left\{i\cdot\log\left(1+\frac{\bar{\rho}^l}{i}\right)\pi_i^l\sum_{\boldsymbol{\delta}_i^{-l}}\left(\Phi^l(\boldsymbol{\delta}_i^{-l})\prod_{m\neq l}\pi_{\delta_m}^m\right)-\frac{\log\pi_i^l}{\theta^l}\right\}, \tag{3.7}$$

where $\boldsymbol{\delta}_i^{-l}=(\delta^1,\ldots,\delta^{l-1},i,\delta^{l+1},\ldots,\delta^L)$ denotes the vector of DoFs with link l having i DoF and $\pi_i^l\sum_{\boldsymbol{\delta}_i^{-l}}\left(\Phi^l(\boldsymbol{\delta}_i^{-l})\prod_{m\neq l}\pi_{\delta_m}^m\right)$ is the long term average frame allocation for link l, when it has i available DoF, considering all possible vectors $\boldsymbol{\delta}_i^{-l}$.

Our objective in static resource allocation is to determine $\Phi^l(\boldsymbol{\delta})$ for all $\boldsymbol{\delta}$ such that total system utility is maximized given the channel distributions and user QoS parameters.

$$\max_{\Phi^l(\boldsymbol{\delta})}\sum_{l=1}^{L}\log[1+E_C^l(\theta^l)]. \tag{3.8}$$

The optimization problem in (3.8) is a non-convex optimization problem due to the min operator in the definition of effective capacity. Hence, we decompose the effective capacity expression into its d states, start to use the auxiliary variable γ^l to denote the effective capacity for each link l, and modify the problem by adding

d additional inequality constraints for each possible DoF for each link.

$$\max_{\Phi^l(\boldsymbol{\delta})} \sum_l \log[1+\gamma^l] \tag{3.9a}$$

$$\gamma^l \leq i \cdot \log\left(1+\frac{\bar{\rho}^l}{i}\right)\pi_i^l \sum_{\boldsymbol{\delta}_i^{-l}} \left(\Phi^l(\boldsymbol{\delta}_i^{-l}) \prod_{m\neq l} \pi_{\delta_m}^m\right) - \frac{\log \pi_i^l}{\theta^l}, \forall l, i \tag{3.9b}$$

$$0 \leq \Phi^l(\boldsymbol{\delta}) \leq 1, \forall l \tag{3.9c}$$

$$\sum_l \Phi^l(\boldsymbol{\delta}) \leq 1, \tag{3.9d}$$

where the first set of constraints in (3.9b) are defined $\forall\, l$ and $i = 1, \ldots, d$. These $L \cdot d$ constraints are obtained by the decomposition of the effective capacity expressions into their states. The constraints in (3.9c) and (3.9d) represent the slot durations of the links and the limited resource constraints, defined for all l and $\boldsymbol{\delta}$.

Note that the above presented problem statement considers all possible $\boldsymbol{\delta}$ and based on these makes the frame allocation decision $\Phi^l(\boldsymbol{\delta})$ for each link l. Since one-shot solution of the presented optimization problem gives all slot allocation decisions for all possible $\boldsymbol{\delta}$, the algorithm is referred to as the static resource allocation in TDMA system. Based on the slot allocation results, the base station reserves each link the optimal length of slot using instantaneous CSI, which gives information on instantaneous $\boldsymbol{\delta}$ per frame.

3.3 Dynamic Resource Allocation Algorithm

Note that the optimization problem stated in the previous section (3.9a) has $L \cdot d^L$ decision variables $\Phi^l(\boldsymbol{\delta})$, and $L \cdot (d+1) + 1$ constraints (i.e. (3.9b), (3.9c) and (3.9d)). As d or L increases, the number of decision variables grow exponentially, which enlarges the search space of the defined problem. In a realistic scenario, e.g. 802.11n, d cannot exceed 4, since the protocol allows a maximum of 4 parallel MIMO streams. However, high L values, i.e. the number of concurrently-active MIMO links or users, need to be taken into consideration, which can potentially introduce large overhead to a realistic system. Thus, we introduce a new algorithm, i.e. dynamic

Table 3.1: Parameters used in dynamic resource allocation algorithm

Parameter	Description
L	total number of MIMO links
l	link index
i	DoF index
d	DoF of MIMO link, i.e. $d = \min\{n_t, n_r\}$
θ^l	QoS of l^{th} link
$\boldsymbol{\xi^k}$	k^{th} vector of DoFs of L links, i.e. $\boldsymbol{\xi^1} = (1, ..., 1)$ and $\boldsymbol{\xi^{d^L}} = (d, ..., d)$
s_k^l	slot allocation for link l for $\boldsymbol{\xi^k}$ in initialization phase
$\delta^l(t)$	DoF of l^{th} link at time t, i.e. $\delta^l(t) = i \in \{1, ..., d\}$
$\widetilde{\boldsymbol{\delta}}_{\boldsymbol{(t)}}$	vector of DoFs of all links, i.e. $\widetilde{\boldsymbol{\delta}}_{\boldsymbol{(t)}} = (\delta^1(t), \ldots, \delta^L(t)) \in \{\boldsymbol{\xi^1}, ..., \boldsymbol{\xi^{d^L}}\}$
$\boldsymbol{\xi}_i^{-l}$	DoF vector with l^{th} link having i DoF, i.e. $\boldsymbol{\xi}_i^{-l} = (\delta^1, \ldots, \delta^{l-1}, i, \delta^{l+1}, \ldots, \delta^L)$
$\widetilde{\phi}^l(\widetilde{\boldsymbol{\delta}}_{\boldsymbol{(t)}})$	instantaneous slot allocation for l^{th} link
$\widetilde{\Phi}^l(\widetilde{\boldsymbol{\delta}}_{\boldsymbol{(t)}})$	updated slot allocation for l^{th} link
π_i^l	probability of l^{th} MIMO link having i DoF
$\bar{\rho}^l$	average transmit SNR of l^{th} MIMO link, i.e. $P{\sigma_{ij}^l}^2/\sigma_n^2$
$\alpha_{(t)}$	weight used in update function
υ^l	auxiliary function $R_i^l - \frac{\log \pi_i^l}{\theta^l}$
$\widetilde{E}_C^l(\theta^l)$	updated effective capacity for link l
$\Psi(t)$	sum of the logarithm of effective capacities, i.e. $\sum_l \log(1 + \widetilde{E}_C^l(\theta^l))$
$\mathbf{0}$	vector of zeros
ε	halt condition

resource allocation in TDMA system, which iteratively solves a simplified version of the optimization problem (3.9a) by updating slot allocations and thus, automatically the effective capacity for each link per frame. This significantly reduces the search space of the optimization problem. However, a modified version of the NUM problem is now solved repetitively.

Due to the changes in the NUM problem statement, the slot allocation variables used before become time-dependent. Thus, frame index t is added to the slot allocation variables. Additionally, they are denoted by $\widetilde{x}$ in order not to confuse the reader with the variables used in static resource allocation. Table 3.1 gives a complete list of the variables used in dynamic resource allocation algorithm.

In dynamic resource allocation, we introduce now the new variable $\widetilde{\phi}^l(\widetilde{\boldsymbol{\delta}}_{\boldsymbol{(t)}})$, the instantaneous frame allocation for link l, which becomes the decision variable of the optimization problem, and we describe $\widetilde{\Phi}^l(\widetilde{\boldsymbol{\delta}}_{\boldsymbol{(t)}})$ as the updated slot allocation for link l based on $\widetilde{\boldsymbol{\delta}}_{\boldsymbol{(t)}}$.

3.3.1 An Optimization Framework for Dynamic Resource Allocation

The natural outcome of dynamically updating the frame allocations is the dynamic update of the effective capacity for each link. Recall that in order to simplify the optimization problem, the derived effective capacity expression, i.e. (3.5), is decomposed into its states, which are added to the set of constraints in the static resource allocation.

In dynamic resource allocation algorithm, we apply the same method with the exception, that in each frame with $\delta^l(t) = i$, the algorithm computes each utility function by taking the logarithm of the auxiliary function $\upsilon^l = R_i^l - \frac{\log \pi_i^l}{\theta^l}$, which is obtained by the decomposition of the effective capacity. With this approach, depending on the instantaneous available DoF of the link, the effective capacity of each link is updated per frame, and the number of constraints obtained by the decomposition reduces from d to 1.

With the mentioned changes, the optimization within each frame with frame index t becomes

$$\max_{\widetilde{\phi}^l(\widetilde{\boldsymbol{\delta}}_{(\boldsymbol{t})})} \sum_l \log\left[1+\upsilon^l\right] \tag{3.10a}$$

$$\upsilon^l \leq i \log_2\left(1+\frac{\bar{\rho}^l}{i}\right)\pi_i^l \sum_{\boldsymbol{\xi}_i^{-l}} \left(\left\{\widetilde{\Phi}^l(\boldsymbol{\xi}_i^{-l}\backslash\widetilde{\boldsymbol{\delta}}_{(\boldsymbol{t})}) + \widetilde{\Phi}^l(\widetilde{\boldsymbol{\delta}}_{(\boldsymbol{t})})\right\}\prod_{m\neq l}\pi_{\delta_m}^m\right) - \frac{\log \pi_i^l}{\theta^l}, \quad \forall\, l \tag{3.10b}$$

$$0 \leq \widetilde{\phi}^l(\widetilde{\boldsymbol{\delta}}_{(\boldsymbol{t})}) \leq 1, \quad \forall\, l \tag{3.10c}$$

$$\sum_l \widetilde{\phi}^l(\widetilde{\boldsymbol{\delta}}_{(\boldsymbol{t})}) \leq 1 \tag{3.10d}$$

where $\widetilde{\Phi}^l(\boldsymbol{\xi}_i^{-l}\backslash\widetilde{\boldsymbol{\delta}}_{(\boldsymbol{t})})$ denotes slot allocations for the link l for all DoF vectors $\boldsymbol{\xi}_i^{-l}$ except $\widetilde{\boldsymbol{\delta}}_{(\boldsymbol{t})}$, for which the slot allocation is updated in frame t. In this form, the optimization problem has a total of L decision variables $\widetilde{\phi}^l(\widetilde{\boldsymbol{\delta}}_{(\boldsymbol{t})})$ and $2L+1$ constraints (i.e. (3.10b), (3.10c), (3.10d)) per frame.

3.3.2 Initialization and Dynamic Control

In the initialization phase of the dynamic resource allocation algorithm, i.e. $t = 0$, each link l is given the a slot length s_k^l for all $\boldsymbol{\xi}^k$ such that $\sum_l s_k^l = 1$. Given this, the optimization problem is solved based on the instantaneous channel characteristics and the given QoS requirements. As a result, $\widetilde{\phi}^l_{(t=0)}$ for all l are obtained.

Next, our algorithm searches for the frame allocation $\widetilde{\Phi}^l$ for each link l that were lastly updated for the vector of DoFs present at t (3.11a) (For $t = 0$, the frame allocations are readily given in the initialization phase). Using a simple linear first order filter (3.11b), $\widetilde{\Phi}^l_{(t)}$ for all l are updated.

Iterations continue until difference of the sum of the logarithm of the updated effective capacities between two consecutive iterations, i.e. $|\Psi_{(t)} - \Psi_{(t-1)}| \leq \varepsilon$, falls below a threshold value ε.

In general, we express the dynamic control part of dynamic resource allocation algorithm in mathematical notation by

$$t^- = \sup_{\tau<t} \left\{ \tau : \widetilde{\boldsymbol{\delta}}_{(\boldsymbol{\tau})} - \widetilde{\boldsymbol{\delta}}_{(\boldsymbol{t})} = \mathbf{0} \right\} \tag{3.11a}$$

$$\widetilde{\Phi}^l(\widetilde{\boldsymbol{\delta}}_{(\boldsymbol{t})}) = \alpha_{(t)} \left[\widetilde{\Phi}^l(\widetilde{\boldsymbol{\delta}}_{(\boldsymbol{t^-})}) \right] + (1 - \alpha_{(t)}) \left[\widetilde{\phi}^l(\widetilde{\boldsymbol{\delta}}_{(\boldsymbol{t})}) \right], \quad \forall l \tag{3.11b}$$

where $\mathbf{0}$ denotes the vector of zeros. (3.11a) looks for the last frame index t^-, for which the vector of DoFs (i.e. $\widetilde{\boldsymbol{\delta}}_{(\boldsymbol{t^-})}$) is the same as the vector of DoFs at frame t, i.e. $\widetilde{\boldsymbol{\delta}}_{(\boldsymbol{t})}$. (3.11b) performs a weighted sum of the instantaneous frame allocation $\widetilde{\phi}^l(\widetilde{\boldsymbol{\delta}}_{(\boldsymbol{t})})$ with the lastly updated frame allocation $\widetilde{\Phi}^l(\widetilde{\boldsymbol{\delta}}_{(\boldsymbol{t^-})})$ for the same vector of DoFs $\widetilde{\boldsymbol{\delta}}$. The weighting term $\alpha_{(t)}$ used in (3.11b) is defined as frame-dependent, which may be useful for certain sets of scenarios. However, throughout our work we prefer to keep it constant at some value $0 \leq \alpha_{(t)} \leq 1$.

3.3.3 Algorithm for Dynamic Resource Allocation

In this section, to understand the dynamic resource allocation in TDMA system, the proposed algorithm is provided below. Table 3.1 displays all the variables used in the algorithm.

Algorithm 1: Dynamic Resource Allocation in TDMA System

Input: $L, d, \bar{\rho}^l, {\sigma_{ij}^l}^2, \sigma_n^2, \alpha_{(t)}, \theta^l, s_k^l, \varepsilon$

for $l = 1$ **to** 1 **do**

  for $k = 1$ **to** d^L **do**

    $\Phi^l(\xi^k) \leftarrow s_k^l$

  end

end

while $|\Psi_{(t)} - \Psi_{(t-1)}| > \varepsilon$ **do**

  $t^- \leftarrow \sup_{\tau<t} \left\{ \tau : \widetilde{\boldsymbol{\delta}}_{(\boldsymbol{\tau})} - \widetilde{\boldsymbol{\delta}}_{(\boldsymbol{t})} = \mathbf{0} \right\}$

  for $l = 1$ **to** L **do**

    $\widetilde{\Phi}^l(\widetilde{\boldsymbol{\delta}}_{(\boldsymbol{t})}) \leftarrow \alpha_{(t)} \left[\widetilde{\Phi}^l(\widetilde{\boldsymbol{\delta}}_{(\boldsymbol{t^-})}) \right] + (1 - \alpha_{(t)}) \left[\widetilde{\phi}^l(\widetilde{\boldsymbol{\delta}}_{(\boldsymbol{t})}) \right]$

  end

  Solve $\max_{\widetilde{\phi}^l(\widetilde{\boldsymbol{\delta}}_{(\boldsymbol{t})})} \sum_l \log\left[1 + \upsilon^l\right]$

  $\upsilon^l \leq i \log_2 \left(1 + \frac{\bar{\rho}_i^l}{i}\right) \pi_i^l \sum_{\boldsymbol{\xi}_i^{-l}} \left(\left\{ \widetilde{\Phi}^l(\boldsymbol{\xi}_i^{-l} \backslash \widetilde{\boldsymbol{\delta}}_{(\boldsymbol{t})}) + \widetilde{\Phi}^l(\widetilde{\boldsymbol{\delta}}_{(\boldsymbol{t})}) \right\} \prod_{m \neq l} \pi_{\delta_m}^m \right) - \frac{\log \pi_i^l}{\theta^l}$

  $0 \leq \widetilde{\phi}^l(\widetilde{\boldsymbol{\delta}}_{(\boldsymbol{t})}) \leq 1$

  $\sum_l \widetilde{\phi}^l(\widetilde{\boldsymbol{\delta}}_{(\boldsymbol{t})}) \leq 1$

  for $l = 1$ **to** L **do**

    Update $\widetilde{E}_C^l(\theta^l)$

  end

  $\Psi(t) \leftarrow \sum_l \log(1 + \widetilde{E}_C^l(\theta^l))$

  $t \leftarrow t + 1$

end

obj $= \max \sum_l \log(1 + \widetilde{E}_C^l(\theta^l))$

3.4 Numerical Results

In this section, we analyze and compare the performance of the two proposed resource allocation methods in TDMA system via numerical experiments. In our experiments, we investigate the behavior of the methods with respect to heterogeneity of users' QoS demands and channel conditions. In this analysis, our aim is to understand how heterogeneous QoS requirements and channel conditions af-

fect the resource allocation decisions in MIMO cellular networks. For this reason, we have not performed large-scale simulations, since having a large number of users would have obscured the effects of varying channel conditions and QoS requirements on resource allocation.

Table 3.2: Resource Allocation in TDMA System Experiments

Experiment #	**Channel Gains**	**QoS Guarantees**
I	$\sigma_{ij}^{1\,2} = \sigma_{ij}^{2\,2} = \sigma_{ij}^{3\,2} = 0.9$	$\theta^1 = \theta^2 = \theta^3 = 0.25$
II	$\sigma_{ij}^{1\,2} = 0.6,\ \sigma_{ij}^{2\,2} = 0.9,\ \sigma_{ij}^{3\,2} = 1.2$	$\theta^1 = \theta^2 = \theta^3 = 0.25$
III	$\sigma_{ij}^{1\,2} = \sigma_{ij}^{2\,2} = \sigma_{ij}^{3\,2} = 0.9$	$\theta^1 = 0.1,\ \theta^2 = 0.25,\ \theta^3 = 0.75$

In our numerical studies, we consider a single cell MIMO network, where $L = 3$ users receive service from a base station, as shown in Figure 3.1. Both the base station and users have three antenna elements, and thus, the maximum degrees of freedom of MIMO links between the base station and users is $d = 3$. The channel gain matrix observed by each receiver l consists of entries $\boldsymbol{H}^l(i,j) = h_{ij}^l \sim \mathcal{CN}(0, \sigma_{ij}^{l\,2})$. Total noise normalized transmit power available at the base station is $P = 10$ Watts. The bandwidth of the wireless channel for each link is set to 1 kHz. The duration of a time slot is one time unit. The users' QoS requirements are indicated by QoS parameter θ^l, $l = 1, 2, 3$.

We performed three experiments for varying channel conditions and QoS parameters. The values of the parameters used in each experiment are depicted in Table 3.2. In the first experiment, we consider homogeneous channel conditions and homogeneous user QoS requirements. In the second experiment, we consider heterogeneous channel conditions but homogeneous QoS requirements. Finally, in the third experiment, we consider heterogeneous QoS requirements and homogeneous channel conditions.

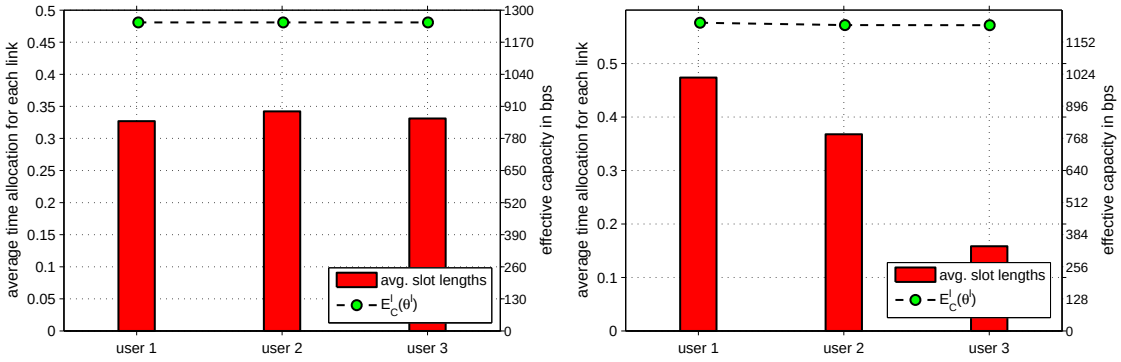
We compare the total utilities obtained in these experiments to equal time allocation case, in which each link transmits one third of the entire frame duration. Hence, with these three experiments we aim to understand how much effect the channel conditions and QoS requirements have on resource allocation. It is worth

emphasizing that the frame allocation decisions presented in the graphs corresponding to the three experiments give information in the average sense, i.e. average frame allocation for each link is obtained by computing the expectation of all the frame allocations dependent on vectors of DoFs with the corresponding DoF probabilities.

Table 3.3: Total Utility and Percentages of Improvement in TDMA System

Experiment #	Equal Time Allocation	static Res. Alloc.	dynamic Res. Alloc.	Change [%]
I	21.0812	21.3954	21.3281	1.49
II	21.0221	21.9185	21.8485	4.26
III	21.0944	22.4313	22.3368	6.33

In Figure 3.4a, we observe that all users are allocated almost the same slot length under static resource allocation, since the channel variance and QoS requirements are the same. However, this is not the case for the dynamic resource allocation (see Figure 3.4b). This is due to the stopping condition $|\Psi_{(t)} - \Psi_{(t-1)}| > \varepsilon$ in the algorithm, which stops iterations until the sum of the utilities falls below a given threshold.

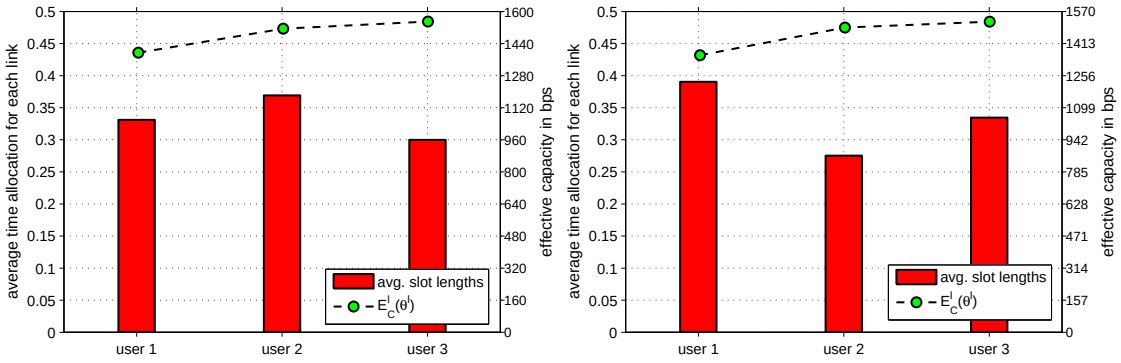


(a) Static Resource Allocation (b) Dynamic Resource Allocation

Figure 3.4: Resource Allocation Decisions for Experiment I.

Despite this, it is worth mentioning that the corresponding effective capacities (see Figure 3.4a and 3.4b) and the total utilities (see Table 3.3) are almost the same. In addition, there is not any significant improvement in total utility compared to the equal time allocation case. Note that the improvement percentages are calculated based on the static resource allocation results.

In Figure 3.5a and in Figure 3.5b, the performance results of static and dynamic resource allocation in Experiment II are given. For static resource allocation, it can be seen that the user with highest channel gain is given the shortest slot length. Despite this, the third user has highest effective capacity. For, dynamic resource allocation, such an observation is not valid. However, both algorithms are again consistent in the effective capacity values, i.e. their total utilities are almost the same as shown in Table 3.3. In this experiment, we see an almost linear relationship between the channel gains and the effective capacity values. One important observation here is that there is a significant 4.26% improvement in our proposed resource allocation scheme compared to the equal time allocation case.

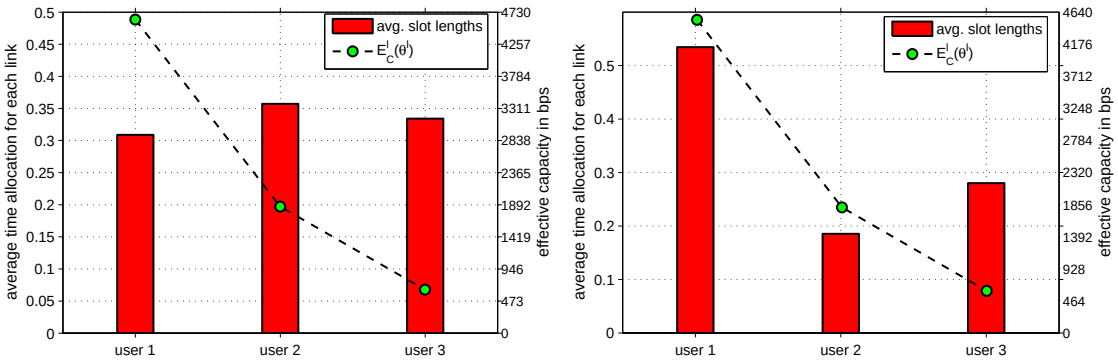


(a) Static Resource Allocation (b) Dynamic Resource Allocation

Figure 3.5: Resource Allocation Decisions for Experiment II.

In Figure 3.6a and In Figure 3.6b, the results of static and dynamic resource allocation of Experiment III are given. The most important observation in this experiment for both the static and the dynamic resource allocations is that the user with the highest QoS-exponent, i.e. the strictest delay-QoS requirement, suffers a lot due to this demand. In Figure 3.6a, despite the fact that user 3 is assigned longer slot than user 1, its effective capacity is much lower. In both Figure 3.6a and 3.6b, it is seen that as the delay-QoS gets stricter, the effective capacity decays more dramatically under the limited resource of unit length frame. Results displayed in Table 3.3 for Experiment III prove the efficiency of our resource allocation algorithm, as the improvement in total utility is 6.33% compared to the equal time allocation.

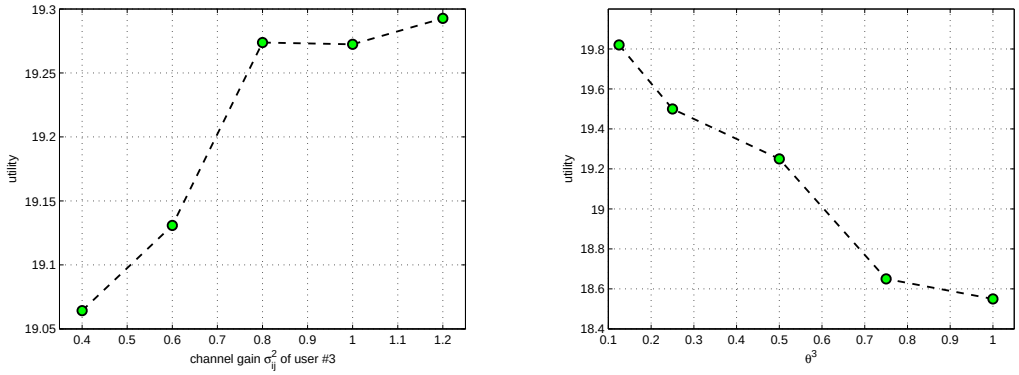
Next, we demonstrate the direct effect of change in one user's parameters, i.e.



(a) Static Resource Allocation (b) Dynamic Resource Allocation

Figure 3.6: Resource Allocation Decisions for Experiment III.

the channel gain ${\sigma_{ij}^l}^2$ and the QoS-exponent θ^l. Data displayed in Figure 3.7a is obtained for $\theta^1 = \theta^2 = \theta^3 = 0.5$, ${\sigma_{ij}^1}^2 = {\sigma_{ij}^2}^2 = 0.5$ and ${\sigma_{ij}^3}^2$ is varied from 0.4 to 1.2. As it is seen, the increase in the channel gain for a single user affects the total utility positively, which is an expected result, as this increases the received power on the receiver side of this user. However, the increase is limited.



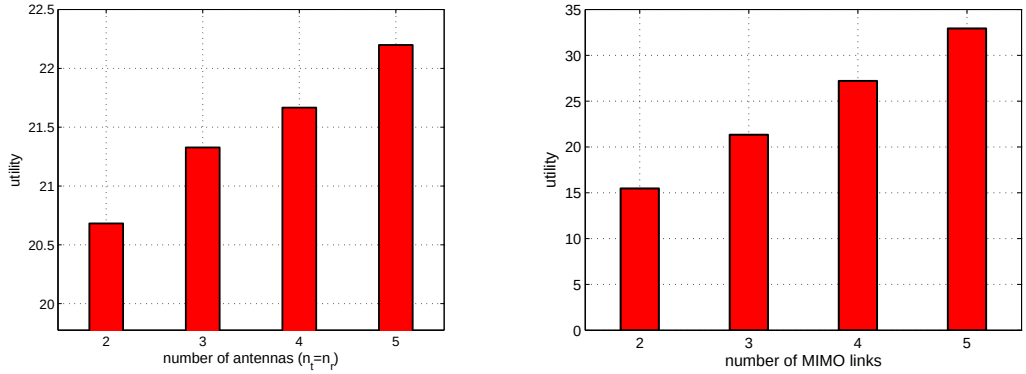
(a) Dependence on Channel Gain (b) Dependence on QoS

Figure 3.7: Utility as a Performance Metric in Frame Allocation I

Data displayed in Figure 3.7b is obtained for ${\sigma_{ij}^1}^2 = {\sigma_{ij}^2}^2 = {\sigma_{ij}^3}^2 = 0.5$, $\theta^1 = 0.25$, $\theta^2 = 0.5$ and θ^3 is varied from 0.125 to 1.0. It is seen that as the QoS demand of third user gets stricter, the total utility rapidly decreases.

Finally, we demonstrate the effect of total umber of antennas per link and the total number of users served by the base station. Figure 3.8a, which is obtained for all three users having the same channel gain ${\sigma_{ij}^l}^2 = 0.9$ and the QoS-exponent

$\theta^l = 0.25$, displays the total utility of the three users when the number antenna elements on both the transmitter and the receiver side of each link is increased from 2 to 5. Even though the transmit power budget at the base station is kept constant, we observe an increase in the total utility. This due to the multiplexing gain introduced by MIMO channels. Data presented in Figure 3.8b, where all users have the same channel gain 0.9 and the QoS-exponent 0.25, shows the change in total utility as the number of users are increased. It is seen that despite the limited resource of unit length frame the total utility increases almost linearly. This behavior is to be explained by the definition of the user utilities given as logarithmic functions of effective capacities.



(a) Dependence on Number of Antennas (b) Dependence on Number of MIMO Links

Figure 3.8: Utility as a Performance Metric in Frame Allocation II

Chapter 4

Resource Allocation in SDMA System

In resource allocation in TDMA system, we have investigated the dynamic allocation of time slot lengths among the users based on the instantaneous CSI feedback acquired from the users. However, in multi-user systems, it is well known that the acquisition of instantaneous CSI introduces significant overhead to system operations. For example, in CDMA/HDR (High Data Rate) system [58], the SNR of each link is measured, from which a value representing the maximum data rate that can be supported is determined. This information is then sent back to the base station via the reverse link Data rate Request Channel (DRC). According to CDMA/HDR specifications, the channel state information is 4 bits long and it is updated every 1.67 ms. If there are 25 users in a cell, 100 bits of channel information has to be sent every 1.67 ms. This requires 60Kbps of channel rate to be dedicated only for CSI feedback. The overhead of acquiring CSI is twice the minimum data rate, and is approximately more than 20% of the average transmission rate as specified by CDMA/HDR specification. Clearly, in a MIMO system, this overhead is expected to be significantly higher. Hence, in this section, we investigate the case when instantaneous CSI is not available at the base station, so the resource allocation is based only on the average channel distributions.

Since the exact instantaneous channel state of the users are not available, dynamic allocation of time slot lengths is in reality not possible. Thus, we determine a

static allocation of resource among the users based on their average channel statistics and QoS requirements. The performance of a static resource allocation policy based on average statistics can never improve upon the performance of an optimal dynamic policy. Hence, by taking interference into consideration we investigate the use of superposition coding in order to simultaneously serve multiple users, and improve the system performance of the proposed static policy.

For this resource problem, we first define a static power allocation algorithm, in which we solve a one-shot NUM problem, with user utilities given as functions of effective capacities of the corresponding MIMO links. In this algorithm, we again use the effective capacity expression we derived in Chapter 3. Next, we define a practical resource allocation for the same problem, in which we adopt a central limit theorem (CLT) approach [59] and use curve fitting for the distributions of the eigenvalues of the MIMO channel gain matrix in finding the effective capacities.

4.1 System Model

We again consider the downlink channel of a single cell in a MIMO based wireless network with a Gaussian broadcast scenario, in which the base station is sending independent messages to L receivers in space division multiple access mode with superposition coding employed for interference cancellation. We assume, the channel gain matrix observed by each receiver l is denoted by $\boldsymbol{H}^l$, consisting of CSCG entries $h_{ij}^l \sim \mathcal{CN}(0, {\sigma_{ij}^l}^2)$. The medium is assumed to have zero-mean Gaussian noise with variance σ_n^2. The base station with full CSI and deployed with n_t antennas to communicate with multiple receivers each with n_r antennas has a total transmit power of P Watts, which is partitioned among multiple links l, i.e. $P\sum_{l=1}^{L}\beta^l = P$, where β^l is the partition of power reserved for link l by the base station. Communication over each MIMO link l with a different QoS requirement θ^l is done simultaneously, i.e. each user is served in the entire frame duration. The base station is assumed to have only full CSI, i.e. it has enough a priori realizations for each link in its memory.

As in the resource allocation in TDMA system, the users obtain a utility function that is a concave function of their effective capacity in both static and the practical

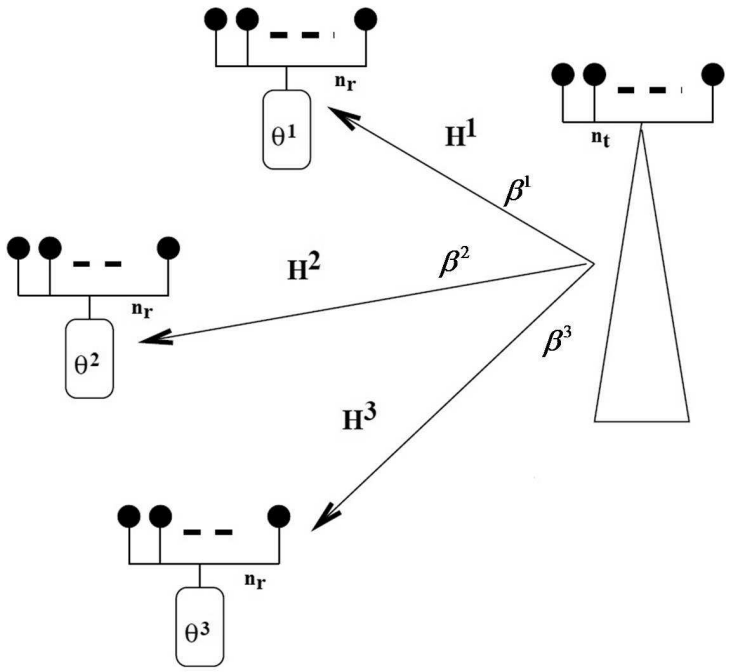


Figure 4.1: Resource allocation in SDMA system with three users.

power allocations. In resource allocation in SDMA system, we again consider a logarithmic utility function that is shown to achieve proportional fairness among the users [53].

4.1.1 Channel Model

Each fading MIMO channel in SDMA system is assumed to be memoryless, i.e. there is no correlation between the channel gain matrix entries h_{ij}^l between two consecutive frames. Each user is served within each frame in the entire frame duration. The channel process is again described by performing state-aggregation, i.e. the process is defined over the available DoF within each frame of length equal to the channel coherence time.

In the static resource allocation algorithm in SDMA system, MIMO channel is again modeled as a discrete Markov chain, where each state i representing the available DoF of a link within a frame occurs with probability π_i. Unlike in TDMA system, the DoF probabilities used to discretize the channel to form Markov model in SDMA system are extracted considering large number of channel realizations, applying only SVD on each channel matrix $\boldsymbol{H}^l$, marking the number of singular values

on the main diagonal of the resulting $\mathbf{\Sigma}^l$ exceeding a sufficiently small threshold level as the available DoF of the link, and counting the number of occurences of each different DoF. The effect of the choice for this threshold value on the effective capacity and the corresponding utility is to be found in Appendix A. So, WF algorithm is not utilized. With the full CSI assumption, the base station knows Markov characterization of each MIMO link.

With the above given procedure to obtain DoF probabilities, the instantaneous channel rate of a channel without interference and with power partition β^l is given by

$$R_i^l = i \cdot \log_2 \left[1 + \frac{{\sigma_{ij}^l}^2 P \beta^l}{\sigma_n^2 \, i} \right] \tag{4.1}$$

with units bps/Hz [54] and SNR is given by $\bar{\rho}^l = \frac{{\sigma_{ij}^l}^2 P \beta^l}{\sigma_n^2}$.

In the existence of interference from neighboring links, when we apply superposition coding, we need to take into account degradation levels ψ^l. In this case the instantaneous channel rate as a function of both β^l and ψ^l is expressed by

$$R_i^l(\beta^l, \psi^l) = i \cdot \log_2 \left[1 + \frac{1}{{\sigma_{ij}^l}^2 P \sum_{k=1}^{L} \mathcal{I}_{\psi^k < \psi^l} \beta^k + \sigma_n^2} \left(\frac{{\sigma_{ij}^l}^2 P \beta^l}{i} \right) \right], \tag{4.2}$$

where $\left({\sigma_{ij}^l}^2 P \sum_{k=1}^{L} \mathcal{I}_{\psi^k < \psi^l} \beta^k + \sigma_n^2 \right)$ is the noise term including the interference from other links for the link l.

4.1.2 Effective Capacity Expression

We adapt (3.5) to static resource allocation in SDMA system algorithm by replacing R_i^l for all $i = 1, \ldots, d$ with (4.2) and expressing them as a function of power partition β^l as

$$
\begin{aligned}
E_C^l(\theta^l, \beta^l, \psi^l) &= \min_{i=0,\ldots,d} \left\{ R_i^l - \frac{\log \pi_i^l}{\theta^l} \right\} \\
&= \min_{i=0,\ldots,d} \left\{ R_i^l(\beta^l, \psi^l) - \frac{\log \pi_i^l}{\theta^l} \right\} \\
&= \min_{i=0,\ldots,d} \left\{ i \cdot \log_2 \left[1 + \frac{1}{{\sigma_{ij}^l}^2 P \sum_{k=1}^{L} \mathcal{I}_{\psi^k < \psi^l} \beta^k + \sigma_n^2} \left(\frac{\sigma_{ij}^l P \beta^l}{i} \right) \right] - \frac{\log \pi_i^l}{\theta^l} \right\}
\end{aligned}
\tag{4.3}
$$

As R_i^l become a function of β^l and ψ^l, we start to denote the effective capacity as a function of θ^l, β^l and ψ^l.

4.2 Static Resource Allocation Algorithm

Our objective in static resource allocation algorithm in SDMA system is to determine the power allocation vector $\boldsymbol{\beta} = (\beta^1, \ldots, \beta^l, \ldots, \beta^L)$ together with the degradation level vector $\boldsymbol{\psi} = (\psi^1, \ldots, \psi^l, \ldots, \psi^L)$ informing about the encoding order of the links such that the total system utility is maximized given the channel characteristics and user QoS requirements given as QoS-exponents.

$$
\max_{\boldsymbol{\beta}, \boldsymbol{\psi}} \sum_{l=1}^{L} \log[1 + E_C^l(\theta^l, \beta^l, \psi^l)]. \tag{4.4}
$$

As the optimization problem in (4.4) is a non-convex optimization problem due to the min operator in the effective capacity expression, the effective capacity expression is decomposed into its d states and the auxiliary variable γ^l is used to denote the effective capacity for each link l in the objective function, i.e. (4.5a). This adds d additional inequality constraints, i.e. (4.5b), per link to the optimization problem.

$$\max_{\boldsymbol{\beta},\boldsymbol{\psi}} \sum_{l=1}^{L} \log[1+\gamma^l] \tag{4.5a}$$

$$\gamma^l \leq i \cdot \log_2 \left[1 + \frac{1}{{\sigma_{ij}^l}^2 P \sum_{k=1}^{L} \mathcal{I}_{\psi^k<\psi^l} \beta^k + \sigma_n^2} \left(\frac{\sigma_{ij}^l P \beta^l}{i}\right)\right] - \frac{\log \pi_i^l}{\theta^l}, \forall l, i \tag{4.5b}$$

$$\sum_{l} \beta^l \leq 1 \tag{4.5c}$$

In case of L active links, one can form $L!$ possible degradation level vectors $\boldsymbol{\psi}$. The above defined optimization problem solves for optimal power allocation vector $\boldsymbol{\beta}$ for each degradation level vector $\boldsymbol{\psi}$, and then decides on the optimal degradation level vector. Given $\boldsymbol{\psi}$, the optimization problem has L decision variables β^l and $L \cdot d + 1$ constraints (i.e. (4.5b) and (4.5c)).

4.3 Practical Resource Allocation Algorithm

In practical resource allocation algorithm, we utilize a CLT approach in finding the effective capacities of the MIMO links. However, this approach requires the extraction of the eigenvalue distributions of a MIMO channel process. In literature, the exact distributions are shown to be too complex to be utilized [60]. Thus, based on the channel characteristics we assume, we apply curve fitting to approximate the distributions.

In this section, we extract the effective capacity of each link as a function of power allocation vector $\boldsymbol{\beta}$. Once this expression is extracted, we solve the optimization problem given in (4.4), in which only $\boldsymbol{\beta}$ becomes the decision variable.

The following subsections introduce the steps taken to form the effective capacity for MIMO channel process. Table 4.1 presents a complete list of parameters used in this section.

Table 4.1: Parameters used in practical resource allocation algorithm

Parameter	Description
L	total number of MIMO links
l	link index
o	encoding order index
d	DoF of MIMO link, i.e. $d = min\{n_t, n_r\}$
P	transmit power budget
θ^l	QoS of l^{th} link
${\sigma_{ij}^l}^2$	variance of the channel entries of l^{th} link
σ_n^2	Gaussian noise present in the medium
$\boldsymbol{\beta}$	power allocation vector, i.e. $\boldsymbol{\beta} = (\beta^1, \ldots, \beta^l, \ldots, \beta^L)$, where β^l is partition of power for l^{th} link
λ_i^l	i^{th} eigenvalue of l^{th} link
$m_i^l({\sigma_{ij}^l}^2)$	mean of i^{th} eigenvalue of l^{th} link as a function of ${\sigma_{ij}^l}^2$
$\bar{m}_i^l$	mean of i^{th} eigenvalue of l^{th} link found solving minimization problem (modeling part)
$v_i^l({\sigma_{ij}^l}^2)$	variance of i^{th} eigenvalue of l^{th} link as a function of ${\sigma_{ij}^l}^2$
$\boldsymbol{\psi}$	degradation level vector, i.e. $\boldsymbol{\psi} = (\psi^1, \ldots, \psi^l, \ldots, \psi^L)$
$\mathcal{I}_{x<y}$	indicator function that takes value 1 when $x < y$
ζ^l	function of power allocation vector $\boldsymbol{\beta}$ used as a parameter, $\zeta^l(\boldsymbol{\beta}) = \frac{d^{-1}P\beta^l}{{\sigma_{ij}^l}^2 \sum_{k=1}^{L} \mathcal{I}_{\psi^k<\psi^l}\beta^k + \sigma_n^2}$
c^o	channel process of o^{th} encoded link, i.e. $c^o = \sum_{i=1}^{d} \log_2 \left[1 + \frac{d^{-1}P\beta^o\lambda_i^o}{{\sigma_{ij}^l}^2 P \sum_{k=1}^{L} \mathcal{I}_{\psi^k<\psi^l}\beta^k + \sigma_n^2}\right] = \sum_{i=1}^{d} \log_2 \left[1 + \zeta^o(\boldsymbol{\beta})\lambda_i^o\right]$
γ^o	effective capacity of a link with encoding order o, i.e. $\gamma^o(\theta^o, \zeta^o(\boldsymbol{\beta})) = E_C^o(\theta^o, \zeta^o(\boldsymbol{\beta})) = \mu_{c^o}(\zeta^o(\boldsymbol{\beta})) - \frac{\theta^o}{2}\sigma_{c^o}^2(\zeta^o(\boldsymbol{\beta}))$
ξ_1^l	parameter characterizing first eigenvalue distribution of link l (exponential distribution)
μ_i^l, σ_i^l	parameters characterizing λ_i^l for $i = 2, \ldots, d$ distribution of link l (log-normal distribution)

4.3.1 Channel Rate Expression

As shown in the algorithms presented so far, the first step in forming an effective capacity expression for a channel process is to express its instantaneous rate. In practical power allocation algorithm, we employ the instantaneous MIMO channel rate formula given in [61], which is the sum of multiple terms, each representing a different parallel and independent data stream or channel, among which transmit power is symmetrically partitioned, i.e. $P\beta^l/d$, with $d = min\{n_t, n_r\}$,

$$R^l = \sum_{i=1}^{d} \log_2 \left[1 + \frac{d^{-1} P \beta^l \lambda_i^l}{{\sigma_{ij}^l}^2 P \sum_{k=1}^{L} \mathcal{I}_{\psi^k < \psi^l} \beta^k + \sigma_n^2} \right], \tag{4.6}$$

where ${\sigma_{ij}^l}^2$ is the variance of the MIMO channel entries, σ_n^2 represents the Gaussian noise present in the medium, and λ_i^l representing the i^{th} eigenvalue of link l, i.e. square of the i^{th} singular value σ_i^{l2} of the MIMO channel gain matrix. The symmetrical partition of the transmit power indicates that WF algorithm is not applied in the practical power allocation algorithm.

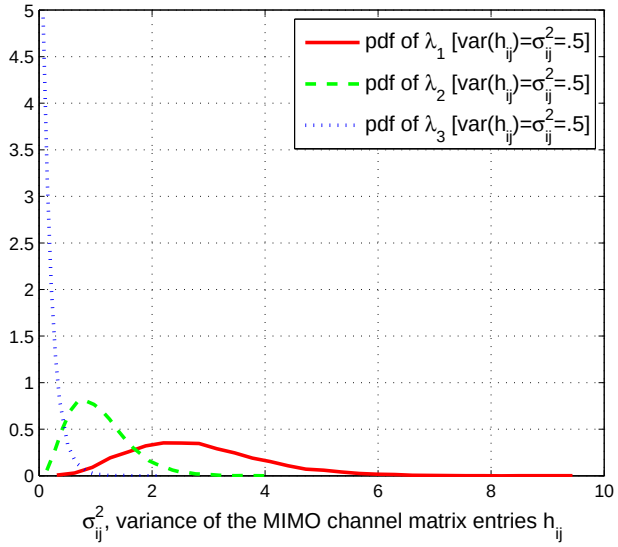


Figure 4.2: Eigenvalue distributions of a MIMO link with $d = 3$

In (4.6), in each of the terms, i.e.

$$\left[1 + \frac{d^{-1} P \beta^l \lambda_i^l}{{\sigma_{ij}^l}^2 P \sum_{k=1}^{L} \mathcal{I}_{\psi^k < \psi^l} \beta^k + \sigma_n^2} \right], \tag{4.7}$$

except λ_i^l, the random variable denoting the i^{th} eigenvalue of link l, all other terms are deterministic. Thus, we denote

$$\zeta^l(\boldsymbol{\beta}) = \frac{d^{-1} P \beta^l}{{\sigma_{ij}^l}^2 P \sum_{k=1}^{L} \mathcal{I}_{\psi^k < \psi^l} \beta^k + \sigma_n^2}, \tag{4.8}$$

and then write, $R^l = \sum_{i=1}^{d} \log_2 \left[1 + \zeta^l(\boldsymbol{\beta}) \lambda_i^l\right]$. Next, we study how λ_i^l's are distributed (Figure 4.2).

4.3.2 Modeling Eigenvalue Distributions of H Matrix

It is well-known that to determine the eigenvalues of $\boldsymbol{H}^l \hat{\boldsymbol{H}}^l$, where $\hat{\boldsymbol{H}}^l$ is the Hermitian transpose of the MIMO channel matrix $\boldsymbol{H}^l$, one needs to solve the characteristic polynomial equation [54]. However, even for a cubic polynomial equation, which corresponds to a MIMO channel with $d = 3$, the formula for roots computation is cumbersome to use. Thus, the polynomial equation is solved approximately by numerical methods. The complexity gets even higher if the aim is to determine the eigenvalue distributions of $\boldsymbol{H}^l$. For this reason, we create approximate eigenvalue distribution models for $\boldsymbol{H}^l$.

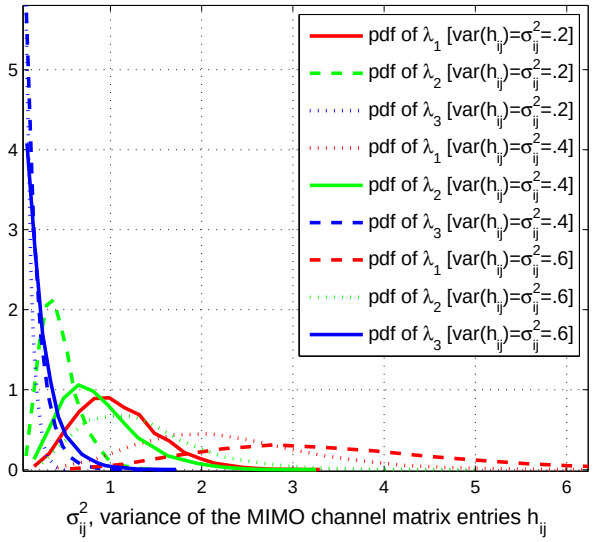


Figure 4.3: Change in eigenvalue distributions of a MIMO link with $d = 3$ as a function of σ_{ij}^2

In Figure 4.2, we give eigenvalue distributions of a single MIMO link with $d = 3$,

where we see that λ_i^l's resemble well-known distributions. In the context of our practical power allocation algorithm, we model λ_1^l as an exponential random variable and other λ_i^l for $i = 2, ..., d$ as log-normal random variables based on the distributions given in Figure 4.2. What affects both the mean and the variance of the λ_i^l's is the variance of the MIMO channel entries, i.e. ${\sigma_{ij}^l}^2$. So, given the channel characteristics, we can a priori extract mean- and variance-functions for each λ_i^l as functions of ${\sigma_{ij}^l}^2$, i.e. $m_i^l({\sigma_{ij}^l}^2)$ and $v_i^l({\sigma_{ij}^l}^2)$, and then extract the parameters characterizing both exponential and log-normal distribution functions, i.e. ξ, μ, σ (Table 4.2). Since each parameter is both eigenvalue- and link-specific, we add both superscript l and subscript i to the parameters, i.e. ξ_i^l, μ_i^l and σ_i^l. Figure 4.3 displays how the distributions change as ${\sigma_{ij}^l}^2$ is increased.

Table 4.2: Eigenvalue Distribution Models

distribution	**pdf**	**mean**	**variance**	**parameters**
Exponential	$\xi e^{-\xi x}$	$\frac{1}{\xi}$	$\frac{1}{\xi^2}$	ξ
Log-normal	$\frac{1}{x\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(lnx-\mu)^2}{2\sigma^2}\right]$	$e^{\mu+\frac{\sigma^2}{2}}$	$(e^{\sigma^2}-1)e^{2\mu+\sigma^2}$	μ, σ

The instantaneous channel rate R^l is in fact the sum of d functions of random variables λ_i^l with $\mathcal{X}_i^l = \log_2\left[1+\zeta^l(\boldsymbol{\beta})\lambda_i^l\right]$, i.e. $R^l = \sum_{i=1}^d \mathcal{X}_i^l = \sum_{i=1}^d \log_2\left[1+\zeta^l(\boldsymbol{\beta})\lambda_i^l\right]$. Now, we characterize $\mathcal{X}_i^l$'s in terms of their means and variances.

λ_1^l assumed to be exponentially distributed, the mean and the variance of its function $\mathcal{X}_1^l$ denoted by $\mu_{\mathcal{X}_1^l}$ and $\sigma^2_{\mathcal{X}_1^l}$ are given as [62], [63],

$$\mu_{\mathcal{X}_1^l} = \log_2(e)\exp\left(\frac{1}{\zeta^l(\boldsymbol{\beta})m_1^l}\right)E_1\left(\frac{1}{\zeta^l(\boldsymbol{\beta})m_1^l}\right), \tag{4.9}$$

$$\sigma^2_{\mathcal{X}_1^l} = \log_2(e)^2\exp\left(\frac{1}{\zeta^l(\boldsymbol{\beta})m_1^l}\right)\Big[\frac{\pi^2}{6}+g^2+2g\log\left(\frac{1}{\zeta^l(\boldsymbol{\beta})m_1^l}\right)+\log^2\left(\frac{1}{\zeta^l(\boldsymbol{\beta})m_1^l}\right)- 2\left(\frac{1}{\zeta^l(\boldsymbol{\beta})m_1^l}\right)\mathcal{F}(\zeta^l(\boldsymbol{\beta}),m_1^l)-\exp\left(\frac{1}{\zeta^l(\boldsymbol{\beta})m_1^l}\right)E_1^2\left(\frac{1}{\zeta^l(\boldsymbol{\beta})m_1^l}\right)\Big], \tag{4.10}$$

where m_1^l is the mean of λ_1^l, g is the Euler constant, E_1 is the exponential integral, $\mathcal{F}(\zeta^l(\boldsymbol{\beta}),m_1^l) = {}_3F_3\left([1,1,1],[2,2,2],-\frac{1}{\zeta^l(\boldsymbol{\beta})m_1^l}\right)$, where ${}_pF_q(\mathbf{n},\mathbf{d},k)$ is the hypergeo-

metric function.

λ_i^l for $i = 2, ..., d$ modeled as log-normally distributed random variables, $\mathcal{X}_i^l$ for $i = 2, ..., d$ with high SNR assumption becomes,

$$\begin{aligned} \mathcal{X}_i^l &= \log_2 \left[1 + \zeta^l(\boldsymbol{\beta})\lambda_i^l\right] \\ &\approx \log_2 \left[\zeta^l(\boldsymbol{\beta})\lambda_i^l\right]. \end{aligned} \tag{4.11}$$

Given this expression, the transformation of the random variable λ_i^l results in $\mathcal{X}_i^l$ being normally distributed, i.e. $\mathcal{X}_i^l \sim \mathcal{N}(\mu_i^l + \zeta^l(\boldsymbol{\beta}), {\sigma_i^l}^2)$. The mean and the variance of λ_i^l for all $i = 2, ..., d$ become $\mu_{\mathcal{X}_i^l} = \mu_i^l + \zeta^l(\boldsymbol{\beta})$ and $\sigma^2_{\mathcal{X}_i^l} = {\sigma_i^l}^2$.

4.3.3 Curve Fitting for Distribution Modeling

In order to be able to use the distribution models mentioned in the previous section, we need to extract the parameters characterizing them, i.e. ξ_i^l, μ_i^l and σ_i^l, for each λ_i^l. However, since eigenvalue distributions do not fully fit to the mentioned distribution models, we apply curve fitting, in which two optimization problems are constructed that try to approximate the mean and the variance of the assumed distribution models given a priori information $m_i^l({\sigma_{ij}^l}^2)$ and $v_i^l({\sigma_{ij}^l}^2)$.

$$\min_{\xi_i^l} \left\{ \left| m_i^l({\sigma_{ij}^l}^2) - \frac{1}{\xi_i^l} \right| + \left| v_i^l({\sigma_{ij}^l}^2) - \frac{1}{{\xi_i^l}^2} \right| \right\} \tag{4.12}$$

$$\min_{\mu_i^l, \sigma_i^l} \left\{ \left| m_i^l({\sigma_{ij}^l}^2) - e^{\mu_i^l + \frac{{\sigma_i^l}^2}{2}} \right| + \left| v_i^l({\sigma_{ij}^l}^2) - (e^{{\sigma_i^l}^2} - 1)e^{2\mu_i^l + {\sigma_i^l}^2} \right| \right\} \tag{4.13}$$

(4.12) is defined for exponentially distributed eigenvalues, i.e. λ_1^l for all l, and (4.13) is formed for log-normally distributed eigenvalues, i.e. λ_i^l for $i = 2, \ldots, d$ and all l. The objective in both optimization problems is to minimize the difference between the given mean and variance of the eigenvalues, and the mean and the variance definitions of the assumed distribution models over ξ_i^l, μ_i^l and σ_i^l.

4.3.4 Effective Capacity Formulation

With no time-correlation among samples, the accumulated channel process for link l, i.e. $C^l(t)$, is simply the addition of t uncorrelated and iid random variables. Expressing the instantaneous channel rate R^l by $c^l(\tau) = \sum_{i=1}^{d} \mathcal{X}_i^l(\tau)$, the cumulative random variable $C^l(t)$ can be expressed by $C^l(t) = \sum_{\tau=0}^{t} c^l(\tau) = \sum_{\tau=0}^{t} \sum_{i=1}^{d} \mathcal{X}_i^l(\tau)$. As $t \to \infty$, the Central Limit Theorem (CLT) can be applied and $C^l(t)$ can be considered as a Gaussian random variable with mean $\mu_{C^l} = t\mu_{c^l}$ and variance $\sigma^2_{C^l} = t\sigma^2_{c^l}$.

The use of this theorem enables us to express both the mean and the variance of the instant channel rate as $\mu_{c^l} = \sum_{i=1}^{d} \mu_{\mathcal{X}_i^l}$ and $\sigma^2_{c^l} = \sum_{i=1}^{d} \sigma^2_{\mathcal{X}_i^l}$, and the statistics of the accumulated channel process as $\mu_{C^l} = t \sum_{i=0}^{d} \mu_{\mathcal{X}_i^l}$ and $\sigma^2_{C^l} = t \sum_{i=0}^{d} \sigma^2_{\mathcal{X}_i^l}$. Note that these statistics are function of $\zeta^l(\boldsymbol{\beta})$, i.e. function of power allocation vector $\boldsymbol{\beta}$ used as a parameter, as shown in the previous section.

Finally, the effective capacity expression for the resulting Gaussian random process $C^l(t)$ is given by [64],

$$E_C^l(\theta^l, \zeta^l(\boldsymbol{\beta})) = \mu_{c^l}(\zeta^l(\boldsymbol{\beta})) - \frac{\theta^l}{2}\sigma^2_{c^l}(\zeta^l(\boldsymbol{\beta})). \tag{4.14}$$

This effective capacity computation method, which we refer to as the CLT approach, is presented in [59], and we study the limits of this approach in [65], where we show that effective capacities computed with this method are valid for relatively loose QoS requirements.

4.3.5 Algorithm for Practical Resource Allocation

In this section, to understand the practical power allocation in SDMA system, the complete algorithm is provided below. Table 4.1 displays all the parameters used in the algorithm. Additional information for the practical encoding order decision is to be found in Appendix B. Given the encoding order decision, the effective capacity of a link with encoding order o becomes only a function of power allocation vector $\boldsymbol{\beta}$ and is denoted by $\gamma^o(\boldsymbol{\beta})$.

Algorithm 2: Practical Resource Allocation in SDMA System

Input: $L, d, P, {\sigma_{ij}^l}^2, \sigma_n^2, m_i^l({\sigma_{ij}^l}^2), v_i^l({\sigma_{ij}^l}^2), \theta^l$

for $l = 1$ **to** L **do**

  $\zeta^l \leftarrow \frac{d^{-1}P}{\sigma_n^2}$

  for $i = 1$ **to** d **do**

    if $i = 1$ **then**

      Solve $\min_{\xi_i^l} \left\{ \left| m_i^l({\sigma_{ij}^l}^2) - \frac{1}{\xi_i^l} \right| + \left| v_i^l({\sigma_{ij}^l}^2) - \frac{1}{{\xi_i^l}^2} \right| \right\}$

      $\bar{m}_1^l \leftarrow (\xi_i^l)^{-1}$

      Compute $\mu_{\mathcal{X}_1^l}(\zeta^l, \bar{m}_1^l)$ and $\sigma^2_{\mathcal{X}_1^l}(\zeta^l, \bar{m}_1^l)$

    else

      Solve $\min_{\mu_i^l, \sigma_i^l} \left\{ \left| m_i^l({\sigma_{ij}^l}^2) - e^{\mu_i^l + \frac{{\sigma_i^l}^2}{2}} \right| + \left| v_i^l({\sigma_{ij}^l}^2) - (e^{{\sigma_i^l}^2} - 1) e^{2\mu_i^l + {\sigma_i^l}^2} \right| \right\}$

      $\mu_{\mathcal{X}_i^l} \leftarrow \left(\mu_i^l + \zeta^l\right)$

      $\sigma^2_{\mathcal{X}_i^l} \leftarrow {\sigma_i^l}^2$

    end

  end

  Compute $\mu_{c^l} = \sum_i \mu_{\mathcal{X}_i^l}$ and $\sigma^2_{c^l} = \sum_i \sigma^2_{\mathcal{X}_i^l}$

  $E_C^l \leftarrow \left(\mu_{c^l} - \frac{\theta^l}{2}\sigma^2_{c^l}\right)$

end

Form $\boldsymbol{\psi}$ based on the set $\left\{E_C^l\right\}_{l=1}^L$

Start using superscript o to identify the links

for $o = 1$ **to** L **do**

  Express $\zeta^o(\boldsymbol{\beta}) = \frac{d^{-1}P\beta^o}{{\sigma_{ij}^o}^2 P \sum_{k=1}^L \mathcal{I}_{\psi^k < \psi^o} \beta^k + \sigma_n^2}$ based on $\boldsymbol{\psi}$

  for $i = 1$ **to** d **do**

    if $i = 1$ **then**

      Express $\mu_{\mathcal{X}_1^o}(\zeta^o, \bar{m}_1^o)$ and $\sigma^2_{\mathcal{X}_1^o}(\zeta^o, \bar{m}_1^o)$

    else

      Express $\mu_{\mathcal{X}_i^o}(\zeta^o)$

    end

  end

  Express $\mu_{c^o}(\zeta^o, \bar{m}_i^o)$ and $\sigma^2_{c^o}(\zeta^o, \bar{m}_i^o)$

  Express $\gamma^o(\boldsymbol{\beta}) = E_c^o(\theta^o, \zeta^o(\boldsymbol{\beta}))$

end

obj $= \max_{\beta^o} \sum_o \log\left[1 + \gamma^o(\boldsymbol{\beta})\right]$

$\sum_o \beta^o \leq 1$

4.4 Numerical Results

In this section, we analyze and compare the performance of the two proposed resource allocation methods in SDMA system in numerical experiments. In our experiments, we investigate the behavior of the methods with respect to heterogeneity of users' QoS demands and channel conditions. In this analysis, our aim is to understand how heterogeneous QoS requirements and channel conditions affect the resource allocation decisions in MIMO networks.

Table 4.3: Resource Allocation in SDMA System Experiments

Experiment #	**Channel Gains**	**QoS Guarantees**
I	$\sigma_{ij}^{1\,2} = \sigma_{ij}^{2\,2} = \sigma_{ij}^{3\,2} = 0.9$	$\theta^1 = \theta^2 = \theta^3 = 0.25$
II	$\sigma_{ij}^{1\,2} = 0.6, \sigma_{ij}^{2\,2} = 0.9, \sigma_{ij}^{3\,2} = 1.2$	$\theta^1 = \theta^2 = \theta^3 = 0.25$
III	$\sigma_{ij}^{1\,2} = \sigma_{ij}^{2\,2} = \sigma_{ij}^{3\,2} = 0.9$	$\theta^1 = 0.1, \theta^2 = 0.25, \theta^3 = 0.75$

Like in the numerical analysis section of resource allocation in TDMA system, we consider a MIMO downlink network where there are $L = 3$ users receiving service from a base station, as shown in Figure 4.1. Both the base station and users have three antenna elements, and thus, the maximum degrees of freedom of MIMO links between the base station and users is $d = 3$. The channel gain matrix observed by each receiver l is denoted by $\boldsymbol{H}^l$, consisting of entries $\boldsymbol{H}^l(i,j) = h_{ij}^l \sim \mathcal{CN}(0, \sigma_{ij}^l)$. Total noise normalized transmit power available at the base station is $P = 10$ Watts. The duration of a time slot is one time unit. The bandwidth of the wireless channel for each link is set to 1 kHz. The users' QoS requirements are indicated by QoS parameter θ^l, $l = 1, 2, 3$.

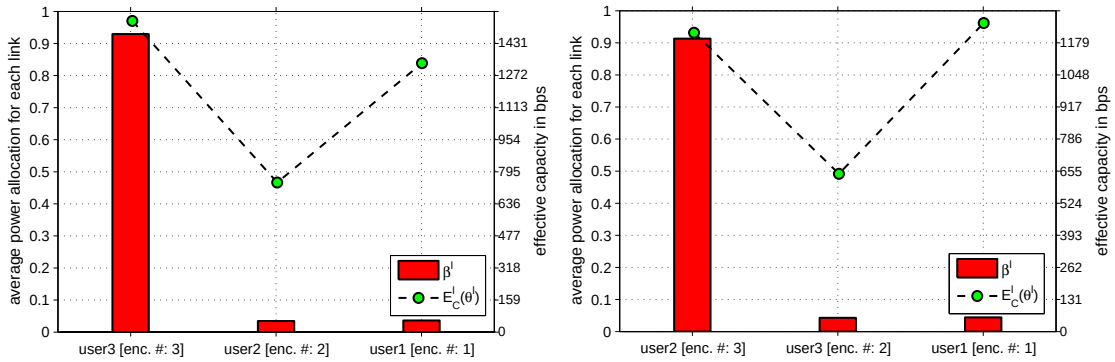
We perform three experiments for varying channel conditions and QoS parameters. The values of the parameters used in each experiment are depicted in Table 4.3. In the first experiment, we consider homogeneous channel conditions and homogeneous user QoS requirements. In the second experiment, we consider heterogeneous channel conditions but homogeneous QoS requirements. Finally, in the third experiment, we consider heterogeneous QoS requirements and homogeneous channel conditions. We compare the total utilities obtained in these experiments to equal

power allocation case for the optimal encoding order of the static resource allocation algorithm, in which each link uses one third of the total transmit power budget of the base station.

Table 4.4: Total Utility and Percentages of Improvement in SDMA System

Experiment #	Equal Power Allocation	static Res. Alloc.	practical Res. Alloc.	Change [%]
I	19.8754	21.1342	20.7076	6.29
II	20.0921	21.2379	21.1358	5.73
III	19.9332	21.1564	20.9220	6.11

In Figure 4.4a and 4.4b, the performance results under Experiment I are given. The first thing to notice is that under the same conditions, the power levels allocated to each user significantly differ from each other.This is due to the fact that each user treats the signals intended for less degraded receivers as additive Gaussian noise. Equal conditions are negatively reflected to the effective capacity values.



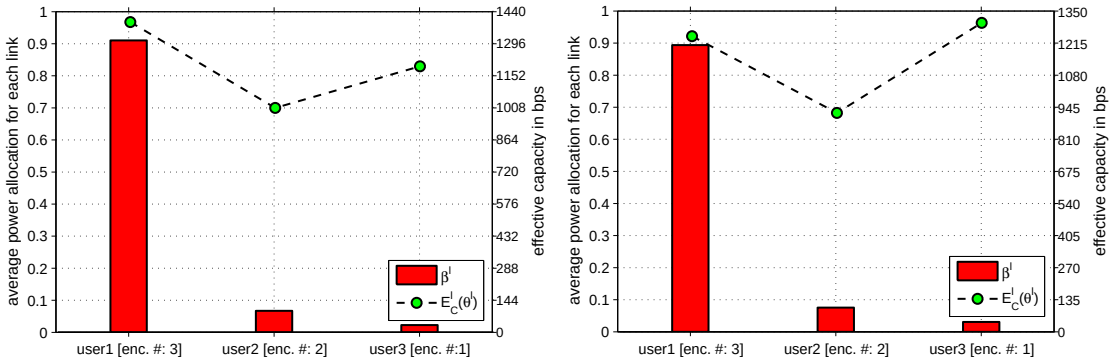
(a) Static Power Allocation

(b) Practical Power Allocation

Figure 4.4: Resource Allocation Decisions for Experiment I.

This means that due to the equal conditions, it is harder for the algorithm to compensate for the large noise. Thus, compared to the results of both Experiment II, the difference between the highest and the lowest effective capacity value is the largest. In addition, we see the encoding order of the links in both figures. The encoding order decisions in both algorithms are different. However, the difference in this decision is not important due to the equal conditions for all links, as any decision would give the same results. An important observation for this experiment

is that our power allocation algorithm improves the total utility by 6.29% compared to the equal power allocation case as shown in Table 4.4. Note that the improvement percentages are calculated based on the static resource allocation results.



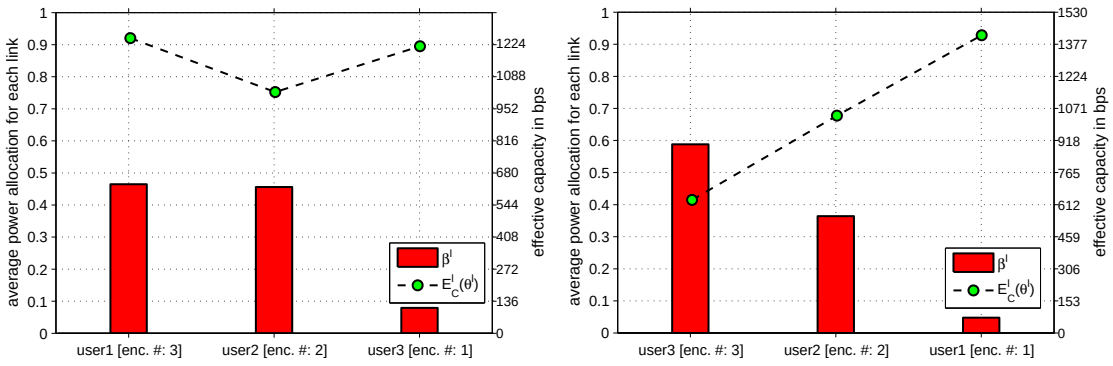
(a) Static Power Allocation

(b) Practical Power Allocation

Figure 4.5: Resource Allocation Decisions for Experiment II.

In Figure 4.5a and 4.5b, we see the results of Experiment II. In general, there is not a significant difference between the results of the static resource allocation and the practical resource allocation. Encoding orders are the same. An important observation is that the user with the highest channel gain is encoded first by the base station, which means that its signal is not interfered by the signals of other users. The user with the worst channel conditions is encoded lastly. However, to compensate for this, the algorithm allocates the largest portion of the transmit power for this user. Due to this compensation, user 1 gets the highest effective capacity. In the practical algorithm, the first encoded user has the largest effective capacity. However, the gap between the effective capacities of the first and last encoded users is small. As data in Table 4.4 shows, our static resource allocation algorithm makes 5.73% improvement in total utility as compared to the equal power allocation case in this experiment.

In Figure 4.6a, which displays the data of Experiment III, we observe that the gap between the power allocations is not as large as the first two experiments. The first encoded user gets the smallest power allocation as expected. This user has the strictest QoS requirement. The last encoded user, i.e. user 1, has a very loose QoS requirement. We normally expect a high effective capacity value for this user.



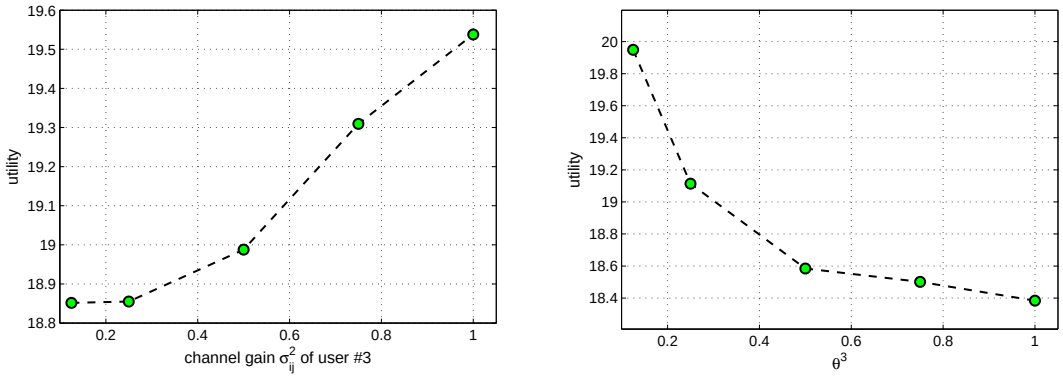
(a) Static Power Allocation

(b) Practical Power Allocation

Figure 4.6: Resource Allocation Decisions for Experiment III.

However, due to interference from other two users, it remains almost at the same level as the first encoded user. It is worth emphasizing that the last encoded user does not need as much power as the last encoded users in the first two experiments due to its very loose QoS demand. Other important observation for this algorithm is that the gap between the highest and the lowest effective capacity values is the lowest as compared to the other experiments. However, Figure 4.6b shows that the power allocations and the encoding order decisions for this experiment are totally different in two algorithms. In Figure 4.6b, it is seen that the user with the loosest QoS demand is encoded first. So, in addition to favoring the loose QoS demand, this user also benefits from the encoding order decision. As a result, its effective capacity is very high. On the other hand, third user with the strictest QoS demand is encoded lastly, which results in a very low effective capacity value despite highest power allocation among the other links. According to Table 4.4, the static resource allocation algorithm makes 6.11% improvement in total utility as compared to the equal power allocation case in this experiment.

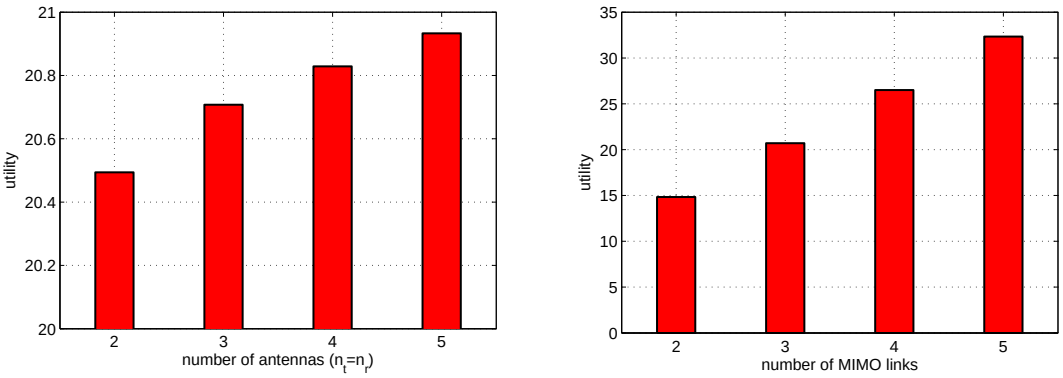
Next, we demonstrate the direct effect of change in one user's parameters, i.e. the channel gain ${\sigma_{ij}^l}^2$ and the QoS-exponent θ^l. Data displayed in Figure 4.7a is obtained for $\theta^1 = \theta^2 = \theta^3 = 0.5$, ${\sigma_{ij}^1}^2 = 0.25$, ${\sigma_{ij}^2}^2 = 0.5$ and ${\sigma_{ij}^3}^2$ is varied from 0.125 to 1.0. As it is seen, with the increase in the channel gain for a single user the total utility grows. This is an expected result, as this increases the received power on the receiver side of this link. However, the increase is limited.



(a) Dependence on Channel Gain (b) Dependence on QoS

Figure 4.7: Utility as a Performance Metric in Power Allocation I

Data displayed in Figure 4.7b is obtained for ${\sigma_{ij}^1}^2 = {\sigma_{ij}^2}^2 = {\sigma_{ij}^3}^2 = 0.5$, $\theta^1 = 0.25$, $\theta^2 = 0.5$ and θ^3 is varied from 0.125 to 1.0. It is seen that the looser the QoS demand of the third user gets, the higher gets the total utility. So, the QoS demand of a single user affects the performance of the entire system.



(a) Dependence on Number of Antennas (b) Dependence on Number of MIMO Links

Figure 4.8: Utility as a Performance Metric in Power Allocation II

Finally, we give the effect of total number of antennas per link and the total number of users served by the base station on the system performance. In Figure 4.8a, which is obtained for all three users having the same channel gain ${\sigma_{ij}^l}^2 = 0.9$ and the QoS-exponent $\theta^l = 0.25$, the total utility of the three users is seen, when the number antenna elements on both the transmitter and the receiver side of each link is increased from 2 to 5. Despite the fact that the transmit power budget at the base station is kept constant, an increase in the total utility is seen. This due to the

multiplexing gain introduced by MIMO technology. Data presented in Figure 4.8b, where all users have the same channel gain 0.9 and the QoS-exponent 0.25, shows the change in total utility as the number of users are increased. As in the resource allocation in TDMA system case, the total utility increases almost linearly despite the limited resource.

Chapter 5

Network Simulations

In this chapter, we show the performance of the resource allocation algorithms presented in Chapter 3 and Chapter 4 in simulation environment. As the simulation environment, we use ns-2 (network simulator 2), which is an open-source discrete event simulator focused on networking research [66].

A discrete event simulator models the operation of a system as a sequence of events in discrete form, i.e. each event occurs in a particular instant and ends with a change of state in the system. Thus, the assumption in all discrete event simulators is that there happens no change in the system between consecutive events.

ns-2 provides basis for research and development in a wide variety of fields such as TCP (transmission control protocol), routing, multicast protocols over both wired and wireless networks. An advantage of ns-2 is that its source codes are available and thus, can be changed such that new protocols can be tested. Another advantage of ns-2 is that due its extensive use there are many modules available on different networking research fields. However, due to its openness it is also subject to bugs that need to be taken care of.

5.1 MIMO Channel Model in ns-2

The most fundamental element in wireless simulation in ns-2 is the node. In order to create a simple wireless communication simulation, one needs first to create two nodes. One node should be defined as the transmitter and the other node needs to

be defined as the receiver, i.e. sink in ns-2. On the transmitting node side, a traffic type (e.g. FTP (file transfer protocol), an analytical arrival process model etc.) and a transmission protocol (e.g. TCP and UDP (user datagram protocol) etc.) need to be configured. Next, there is need to define the PHY, MAC and Link layer properties of the wireless link between the nodes. Finally, the desired simulation can be performed according to the settings written on the OTCl code.

As given in the previous chapters, this thesis focuses on resource allocation policies among multiple links with antenna arrays on both the transmitter and the receiver side. However, ns-2 supports one antenna per node. An official IEEE 802.11n protocol module, i.e. wireless networking standard using multiple antennas to increase data rates, is not present in ns-2.

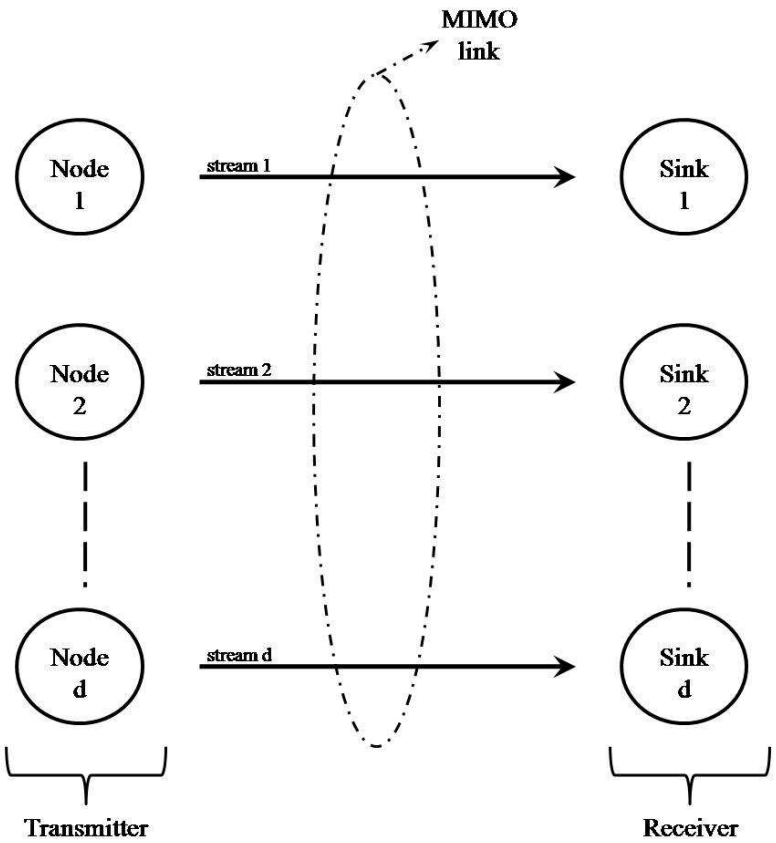


Figure 5.1: ns2 channel model.

As a result, we decided to model each node as an antenna element on both the transmitter and the receiver sides. Next, we decomposed each MIMO channel into independent parallel data streams. Each particular data stream is modeled as a data transmission between two particular nodes creating a pair, each of which is independent of the remaining pairs of the same MIMO link. In Figure 5.1, we give the representative MIMO channel model with d antenna elements on both sides of the link.

Next, we created a module, which proportionally allocates the limited power of the base station, i.e. transmitter side of each link, based on the instantaneous eigenvalues of the MIMO channel gain matrix. ns-2 originally does not consider random variations in the wireless channel. Instead, ns-2 focuses on the reduction in power density of the electromagnetic wave as it propagates through space, i.e. path loss. Since we do not consider signal attenuation, we have disabled path loss in ns-2 and integrated fading to the power module to simulate random variations in the wireless channel.

5.2 System Model

In the ns-2 simulations, we again adopt the MIMO based downlink scenario. The base station deployed with $n_t = 3$ antennas serves a total of 4 users with $n_r = 3$. The base station is assumed to have both full and instantaneous CSI for the TDMA case and only full CSI for the SDMA case. The assumptions for the channel gain matrix entries and noise distributions given in Chapter 3 and 4 are also valid in this chapter.

Table 5.1: PHY Layer Parameters

Parameter	**Value**
Carrier Frequency	2.4 GHz
Overall Channel Bandwidth	22 MHz
Overall Transmit Power	10 dBm
Overall Noise Power	-20 dBm

In Table 5.1, PHY layer parameters used in ns-2 are given. [67] is taken as a

reference in selecting these values. The resource allocation algorithms under investigation are integrated to Mac/802.11 module in ns-2.

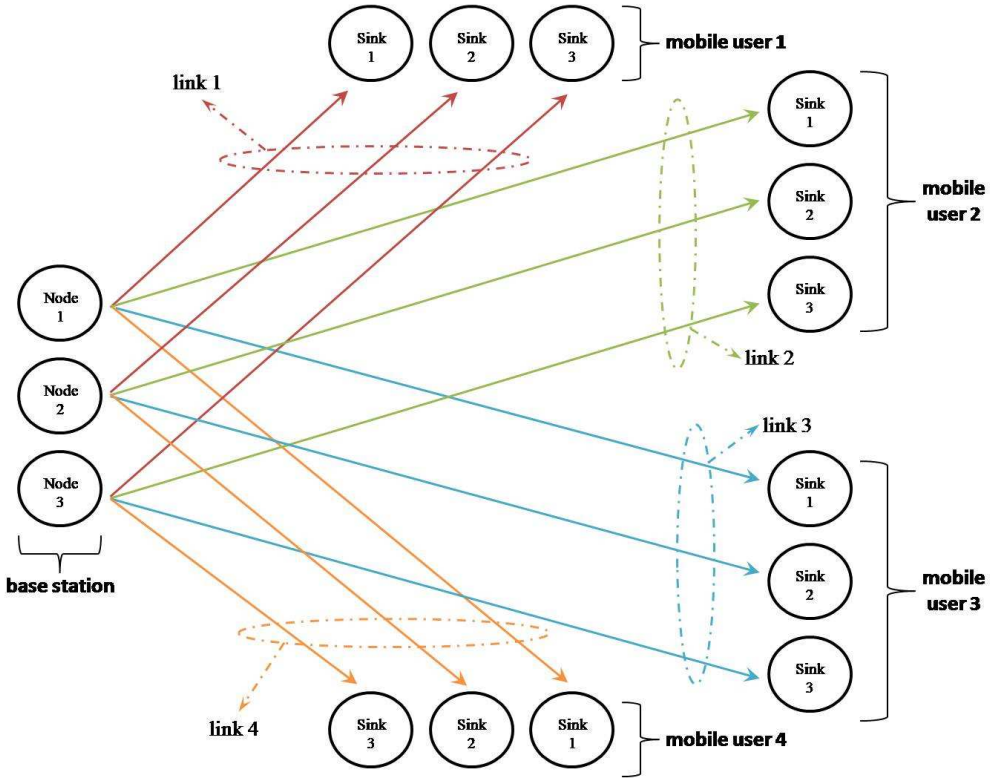


Figure 5.2: ns-2 simulation setting.

In Figure 5.2, the overall setting used in ns-2 is given. Since path loss is disabled, the distance between the nodes are to be neglected. Only, the random channel variations affect the resource allocation policies. Additionally, the users are stationary.

5.2.1 Model for Wireless Channel Interference

In discrete event simulators, modeling the wireless channel interference is crucial, since the accuracy of the simulation data is determined by the level of detail of the adopted model. In a network simulator, a successful data packet reception occurs

only if the power of the received packet exceeds given thresholds, where interference in the wireless channel can be integrated.

ns-2 uses the so-called capture threshold model, which only accounts for one interferer at a time. Since this is not applicable to our simulation setting for SDMA, where interference is taken into consideration due to simultaneous access of multiple links to the wireless medium, we adopt the additive interference model (SINR thresholding) in [67] by omitting the carrier sense threshold, which checks the idleness of the wireless medium prior to transmission to avoid packet collisions, since in our resource allocation algorithm in SDMA system, all users are forced to transmit the entire frame duration.

In our wireless channel interference model, each receiver checks only for the ratio of the intended received signal to the sum of the noise in the medium and the other signal powers, i.e. signal to interference and noise ratio (SINR). The superposition coding uses this interference model.

5.2.2 Actual Delay Computation

In Chapter 2, the delay violation of data sent by a source with constant rate μ is given in (2.10). From the effective capacity perspective, we know that once the effective capacity of a channel process is extracted as a function of θ, i.e. $E_C(\theta)$, given a θ^*, μ can as high as $E_C(\theta^*)$, such that the target delay D_t satisfies ϵ. Assuming that the probability of the queue not being empty is 1, i.e. $\gamma(\mu) = 1$, (2.10) can be rewritten as

$$\epsilon = e^{-\theta^* E_C(\theta^*).D_t}. \tag{5.1}$$

Next, given a desired delay violation probability ϵ, the target delay corresponding to both ϵ and θ^* can be computed by

$$D_t = \frac{\log(1/\epsilon)}{\theta^* E_C(\theta^*)}\,. \tag{5.2}$$

It is important to note that (5.2) does not consider data transmission in packet format, i.e. its direct use would result in delay value for bit-by-bit transmission of data. However, data is transmitted packet-by-packet in real systems, which is also the case in discrete event network simulators. Hence, (5.2) is modified as

$$D_t = \frac{\log(1/\epsilon)\, p_{length}}{\theta^* E_C(\theta^*)}, \tag{5.3}$$

where p_{length} denotes the packet length in bits. (5.3) considers packet transmission rate, as the effective capacity is normalized by the packet length.

5.2.3 Delay Analysis in ns-2

In order to analyze the output produced by ns-2, the actual delay computation is necessary. However, this requires the prior numerical computation of the effective capacity values by the proposed algorithms. Thus, the target delay D_t of each link is numerically computed before the ns-2 simulations given ϵ, which is kept at the same value for all links throughout our simulations.

In ns-2 simulations, an input traffic with a constant rate is defined for each transmitting node based on the numerically computed effective capacity value. The simulations are run for relatively long time durations. In the analysis of the resulting output files, the delay value corresponding to the delay violation probability, i.e. ϵ, is extracted for each link. Once the delay value for a link is obtained, the corresponding throughput, i.e. effective capacity, is calculated by counting the number of successfully received packets not violating the obtained delay value.

In the following section, we first give the delay values obtained numerically, compare them to the ones obtained by the analysis of ns-2 data, and then give figures displaying the resource allocation decisions and the corresponding effective capacity values extracted in ns-2.

5.3 Simulation Results

In this section, we analyze the delay performance and demonstrate the resource allocation decisions in TDMA and SDMA systems in simulation environment. For the resource allocation in TDMA system case, we employ the dynamic algorithm. For the resource allocation in SDMA case, we utilize the practical algorithm.

Table 5.2: ns-2 Resource Allocation Experiments

Experiment #	**Channel Gains**	**QoS Guarantees**
I	$\sigma_{ij}^{1\,2} = \sigma_{ij}^{2\,2} = \sigma_{ij}^{3\,2} = \sigma_{ij}^{4\,2} = 0.9$	$\theta^1 = \theta^2 = \theta^3 = \theta^4 = 0.25$
II	$\sigma_{ij}^{1\,2} = 0.6$, $\sigma_{ij}^{2\,2} = 0.8$, $\sigma_{ij}^{3\,2} = 1.0$, $\sigma_{ij}^{4\,2} = 1.2$	$\theta^1 = \theta^2 = \theta^3 = \theta^4 = 0.25$
III	$\sigma_{ij}^{1\,2} = \sigma_{ij}^{2\,2} = \sigma_{ij}^{3\,2} = \sigma_{ij}^{4\,2} = 0.9$	$\theta^1 = 0.1$, $\theta^2 = 0.25$, $\theta^3 = 0.5$, $\theta^4 = 0.75$

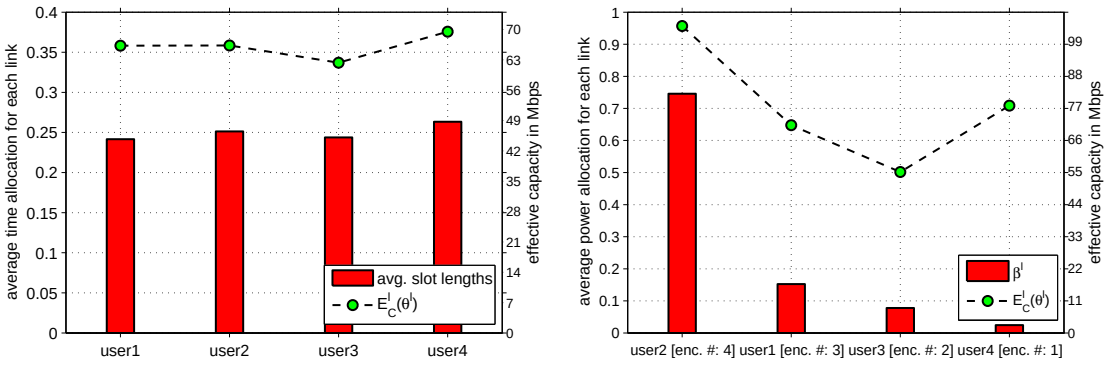
We perform three experiments for varying channel conditions and QoS parameters. The values of the parameters used in each experiment are depicted in Table 5.2. In the first experiment, we consider homogeneous channel conditions and homogeneous user QoS requirements. In the second experiment, we consider heterogeneous channel conditions but homogeneous QoS requirements. In the third experiment, we consider heterogeneous QoS requirements and homogeneous channel conditions. In all experiments, p_{length} is set to 1024 *bytes* and ϵ is taken as 0.001 for all links.

Before presenting the results, we give the Shannon capacity limit and the maximum achievable effective capacity for each experiment in Table 5.3 to be used as a reference to comment on the simulation data. The capacity limits in Table 5.3 are computed by selecting the user with the best channel gain and the loosest QoS

Table 5.3: Capacity Limits

Experiment #	**Shannon Capacity Limit**	**Maximum Effective Capacity**
I	569.55 Mbps	387.95 Mbps
II	596.47 Mbps	406.17 Mbps
III	569.55 Mbps	398.25 Mbps

demand in each experiment, and allocating all the resources, i.e. frame length and the power, to this user.



(a) TDMA system

(b) SDMA system

Figure 5.3: Performance of Resource Allocation Algorithms for Experiment I in ns-2.

5.3.1 Experiment I

For Experiment I, the numerical solution of the NUM problem results in effective capacity values 72.64 $Mbps$, 72.09 $Mbps$, 66.57 $Mbps$ and 74.56 $Mbps$ in TDMA system, and 76.05 $Mbps$, 113.28 $Mbps$, 59.21 $Mbps$ and 83.27 $Mbps$ in SDMA system corresponding to the users 1 through 4, respectively. Evaluation of these numerical effective capacity values in (5.3), results in actual delay values 3.1 ms, 3.1 ms, 3.4 ms and 3.0 ms in dynamic algorithm proposed for TDMA system, and 3.0 ms, 2.0 ms, 3.8 ms and 2.7 ms in practical algorithm proposed for SDMA system for the users 1 through 4, respectively.

Table 5.4: Delay Values for Users in Experiment I

	TDMA System				**SDMA System**			
user #	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**
numerical result in ms	3.1	3.1	3.4	3.0	3.0	2.0	3.8	2.7
simulation result in ms	3.7	3.6	3.5	3.5	3.1	2.7	3.8	3.1

Analysis of ns-2 data resulted in the delay values given in Table 5.4. An initial observation for both algorithms is that the delay values obtained in simulation environment are higher. Based on (5.3), this behavior can be explained by lower effective capacity values obtained in ns-2 based on the extracted delay values matching ϵ as shown in Figure 5.3a and Figure 5.3b. A deeper analysis on the receiver side of each link shows that there are relatively high number of packets dropped at the receiver

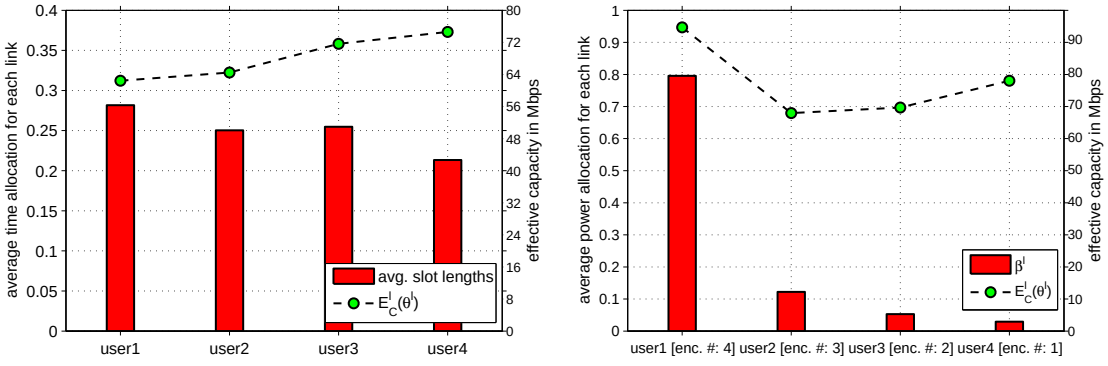
side. Dropping of the packets can be explained by the transmitter trying to send packets at a much higher rate than the channel can support. As a result of this, the power consumed per packet transmission decreases, and due to the receiver sensitivity, i.e. power threshold, which "accepts" packets based on their power, some packets are dropped. This threshold in fact is an indicator for the ability of the receiver to successfully decode a packet.

In Figure 5.3a and 5.3b, the effective capacity values together with resource allocation decisions are given in detail. The difference between the effective capacity values obtained in numerical experiments and simulation environment is between $6.8 - 9.6\%$ in TDMA system, and $7.7 - 8.6\%$ in SDMA system. We see that the sum of the effective capacities of four users are lower than the limits given for this experiment in Table 5.4. This shows that the use of the system by multiple users makes it operate at a much lower capacity.

An important observation for the delay values in TDMA system is that they are relatively in the same range. This observation is also valid for the resource allocation decisions. It is seen that each user is given almost the same length of slot. As a result, the effective capacities of the users are very close. This is due to the equivalent conditions for each user. In Figure 5.3b, we observe that the effective capacity values are different. As a result of this, their corresponding delay values also differ. Furthermore, the encoding order decision is given as *link 4, link 3, link1, link 2* by the practical power allocation algorithm. In fact, under the same conditions all 4! possible encoding order decisions are equivalent. Furthermore, it is again seen that the last encoded user is allocated the largest portion of power resulting in the highest effective capacity among the other users despite the interference from all other links. In comparison to the third and first links, second link encoded firstly has a higher effective capacity despite relatively low power allocation. This shows the impact of the interference on the effective capacity of the links. As a result of diverse effective capacity values obtained in SDMA, the actual delay values are different.

5.3.2 Experiment II

For Experiment II, the numerical solution of the utility maximization problem gives effective capacity values 67.80 $Mbps$, 68.48 $Mbps$, 77.24 $Mbps$ and 81.11 $Mbps$ in TDMA system, and 102.72 $Mbps$, 73.13 $Mbps$, 73.68 $Mbps$ and 84.04 $Mbps$ in SDMA system corresponding to the users 1 through 4, respectively. Evaluating these numerical effective capacity values in (5.3), results in actual delay values 3.3ms, 3.3 ms, 2.9 ms and 2.8 ms in dynamic algorithm proposed for TDMA system, and 2.2 ms, 3.1 ms, 3.1 ms and 2.7 ms in practical algorithm proposed for SDMA system for the users 1 through 4, respectively.



(a) TDMA system (b) SDMA system

Figure 5.4: Performance of Resource Allocation Algorithms for Experiment II in ns-2.

We give delay analysis results of ns-2 data in Table 5.5. The numerically found delay values are again lower than the ones found in simulation environment. In general, it can be stated that simulation data follows the trend in the numerically computed delay values, i.e. as the effective capacity decreases the delay of the packets get higher. So, the simulation data proves the relation given in (5.3). Due to the same QoS requirements in form of QoS-exponent, we observe that the actual delay values are almost inversely proportional to the effective capacity values. The same observation is also true for the first experiment.

The difference between the effective capacity values obtained in numerical experiments and simulation environment in this experiment is between $6.2 - 8.7\%$ in TDMA system, and $6.3 - 9.7\%$ in SDMA system. It is again seen that the sum of the

Table 5.5: Delay Values for Users in Experiment II

	TDMA System				SDMA System			
user #	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**
numerical result in ms	3.3	3.3	2.9	2.8	2.2	3.1	3.1	2.7
simulation result in ms	3.7	3.6	3.2	3.3	2.4	3.7	3.6	3.1

effective capacities of four users are lower than the limits given for this experiment in Table 5.4.

In Figure 5.4a and 5.4b, we see the effective capacities and the resource allocation results of Experiment II in ns-2. Important observations for data displayed on 5.4a is that the link with highest channel gain is assigned the shortest average length of slot, and there is almost a linear relationship between the channel gain and the effective capacity of a link, which are also valid for the results given in the second experiment in Chapter 3. The most important observation in Figure 5.4b is that the encoding order decision is given as *link 4, link 3, link 2, link 1*, which indicates that the user with the highest channel gain is encoded firstly, i.e. receives no interference, and the user with the lowest channel gain is encoded lastly.

5.3.3 Experiment III

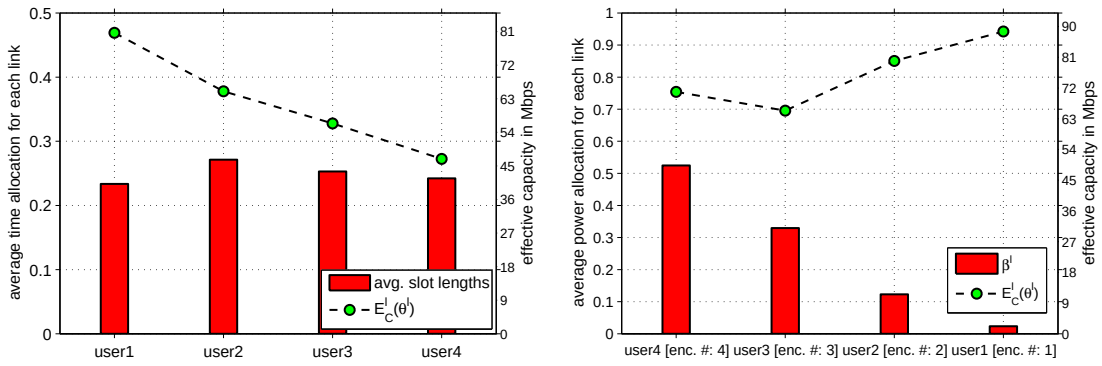
In Experiment III, the effective capacity values obtained by numerical solution of the defined maximization problem are 86.79 $Mbps$, 70.34 $Mbps$, 61.90 $Mbps$ and 50.86 $Mbps$ in TDMA system, and 95.05 $Mbps$, 86.19 $Mbps$, 70.93 $Mbps$ and 75.56 $Mbps$ in SDMA system corresponding to the users 1 through 4, respectively. Actual delay values obtained based on these numerical effective capacity values using (5.3) results in actual delay values $6.5ms$, $3.2ms$, $1.8ms$ and $1.5ms$ in dynamic algorithm proposed for TDMA system, and $0.8ms$, $1.3ms$, $3.2ms$ and $7.5ms$ in practical algorithm proposed for SDMA system for the users 1 through 4, respectively.

Table 5.6 presents delay values obtained as a result of the analysis of ns-2 data together with the delay values computed numerically. An important observation for the actual delay values is that the difference among the user delays is the highest in this experiment compared to the first two experiment. This to be explained by the

Table 5.6: Delay Values for Users in Experiment III

	TDMA System				SDMA System			
user #	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**
numerical result in ms	6.5	3.2	1.8	1.5	0.8	1.3	3.2	7.5
simulation result in ms	7.7	3.6	2.5	1.8	1.1	1.7	3.2	7.1

diverse QoS requirements of the users in form of QoS-exponent directly affecting the delay value in (5.3). Thus, we do not observe an inverse proportionality between the effective capacity value and the actual delay value for each user. It is again important to note that the simulation data tends to follow the numerically found delay values.



(a) TDMA system

(b) SDMA system

Figure 5.5: Performance of Resource Allocation Decisions for Experiment III in ns-2.

In Figure 5.5a and 5.5b, the simulation data of Experiment *III* is presented. In this experiment, the difference between the effective capacity values obtained in numerical experiments and simulation environment in this experiment is between $7.6 - 9.8\%$ in TDMA system, and $6.6 - 8.5\%$ in SDMA system. In this experiment, the capacity limits given in Table 5.3 are again not reached.

For the resource allocation in TDMA system, it is seen that as the QoS requirement gets stricter, the effective capacity decreases significantly, which cannot be compensated by the resource allocation algorithm. The user with the loosest QoS constraint is given the shortest slot length. However, it does not differ much from the slot length allocations to other users. An important observation in Figure 5.5b

is that the encoding order decision is given as *link 1, link 2, link 3, link 4*, which indicates that the user with the loosest QoS demand is encoded firstly, i.e. receives no interference, and the user with the strictest QoS demand is encoded lastly. Similar to the results of the dynamic resource allocation algorithm in TDMA system, it can be stated for this experiment that as the QoS demand gets looser, the effective capacity increases. However, this statement cannot be generalized. A final observation for the resource allocation decisions in SDMA system is that the gap between the highest and the lowest effective capacity values is at its lowest value among all three experiments. This can be explained by the algorithm trying to compensate for the large variance in QoS demands by allocating the limited power more evenly rather than concentrating the largest portion at one user.

Chapter 6

Conclusion and Future Work

In this thesis, we propose two main resource allocation approaches with delay-QoS provisioning in MIMO based downlink systems, i.e. resource allocation in TDMA system and resource allocation in SDMA system. For this, applying state aggregation we first formulate the effective capacity of a general MIMO process by utilizing effective capacity theory, which converts average delay constraints into equivalent average rate constraints. Delay-aware resource control is then performed by forming a NUM problem, where a logarithmic utility function of effective capacity is defined for each user, which is shown to achieve proportional fairness among the users.

For the resource allocation in TDMA system, we present a static and a dynamic resource allocation algorithm, where optimal length of slot for each link is allocated based on the channel conditions and the QoS requirements. In both of these algorithms, we assume both full and instantaneous CSI at the base station. In static resource allocation algorithm, a one-shot optimization problem is solved by taking into account all possible DoF vectors of the links based on full CSI, whose result is to be applied based on the instantaneous channel conditions, i.e. available DoFs, with instantaneous CSI knowledge. In dynamic TDMA, we only assume that the prior knowledge of the state probabilities of all the links is available at the base station, and iteratively perform slot length allocation among the links based on instantaneous CSI.

For the resource allocation in SDMA system, considering interference we present a static resource allocation algorithm and a practical resource allocation algorithm,

where optimal portion of the limited power available at the base station is allocated to simultaneously transmitting links based on the overall channel characteristics and the QoS requirements in the long term. To deal with interference, we utilize superposition coding. In static resource allocation algorithm, we form an optimization problem similar to the one formed in static resource allocation algorithm in TDMA system by adopting the extracted effective capacity expression for the MIMO channel process. In practical resource allocation algorithm, we first extract the eigenvalue distributions of the MIMO channel process as a function of the MIMO channel gain entries' statistics by applying curve fitting. Next, applying CLT and considering each cumulative channel process as a Gaussian process, we extract a simple effective capacity expression, based on which we form another optimization problem.

After introducing the algorithms, we show their performance via numerical experiments. Finally, we demonstrate the performance of the algorithms together with actual delay analysis in simulation environment by considering realistic channel models.

As a future work, the effective capacity expression for the MIMO channel process, which we believe is the most significant contribution of this thesis, may be integrated into more sophisticated networking research fields such as multi-hop wireless networks and routing protocols exploiting MIMO technology, which particularly lack such an expression that takes into account QoS considerations. In particular, we believe that the integration of this expression into the NUM problem given in [9] may be a real contribution to wireless networking research. As another future work, we have plans to run the proposed algorithms in a more advanced and up-to-date discrete event simulator, which preferably has a ready-to-use IEEE 802.11n module.

Appendix A

DoF Probability Computation in SDMA System

In Chapter 4, it is stated that the DoF probabilities required for the discretization of the MIMO channel process are found by applying only SVD on each channel matrix $\boldsymbol{H}^l$, marking the number of singular values on the main diagonal of the resulting $\boldsymbol{\Sigma}^l$ exceeding a sufficiently small threshold level as the available DoF of the link, and counting the the number of occurences of each different DoF. As it should be noticed, the choice of this threshold value may affect the effective capacity of a MIMO link.

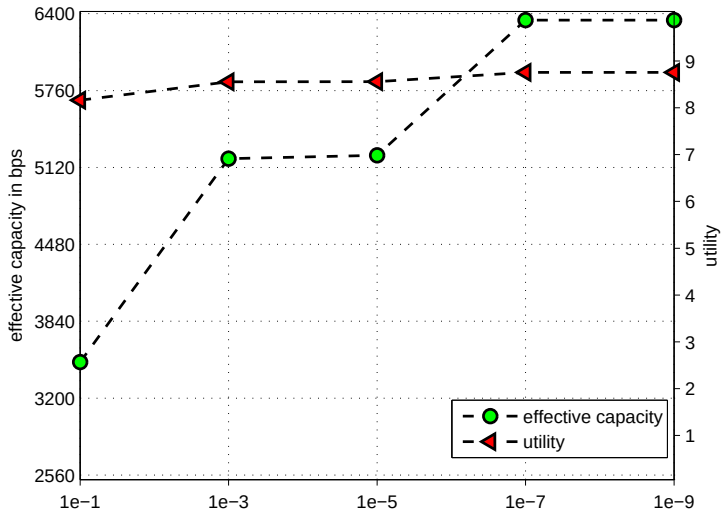


Figure A.1: Effect of Choosing the Threshold Value in Finding DoF Probabilities

As a result of the choice, the utility of the link may take different values, which may

change the resource allocation decisions in the NUM problems. Hence, we study in this part of the thesis the effect of the choice for this threshold value.

Figure A.1 displays both the effective capacity and the corresponding utility of a single MIMO link with 3 antennas on both the receiver and the transmitter side. The MIMO link with gain $\sigma_{ij}^2 = 1.0$ is given a bandwidth of 1kHz. The noise-normalized power of the transmitter is 10W, and the link has a QoS-exponent $\theta = 0.25$. The threshold level is varied from 10^{-1} to 10^{-9}. The data displayed is obtained by taking 1 million channel realizations.

The first thing to notice in Figure A.1 is that as the threshold value is lowered the effective capacity increases. This is to be explained by π_3 getting a higher value due to the decreased threshold. For threshold level 10^{-9}, the DoF probabilities are found as $\pi_1 = 0$, $\pi_2 = 0.0001$ and $\pi_3 = 0.9999$. On the other hand, for the threshold being equal to 10^{-1}, the probabilities are found as $\pi_1 = 0.0001$, $\pi_2 = 0.2595$ and $\pi_3 = 0.7404$. As a result of the lower probability for the MIMO link having 3 DoF, its effective capacity value decreases dramatically.

However, the change in the utility values is not as large as it is the case for the effective capacity values. This is due to the use of the logarithmic utility function shown to provide proportional fairness. Throughout the simulations presented in this thesis, the threshold is taken as 10^{-3}.

Appendix B

Decision of Encoding Order

For the resource allocation in SDMA system the determination of the encoding order of the links is very important, because the given optimization problem for the static power allocation is solved $L!$ times, since there are $L!$ possible degradation level vectors $\boldsymbol{\psi}$.

For the practical power allocation algorithm presented in Chapter 4, we examined the effects of the QoS-exponents, i.e. θ^l, and the variance of the MIMO channel matrix entries, i.e. ${\sigma_{ij}^l}^2$, on the encoding order of the links to reduce the search space of the optimization problem. For this, we performed controlled experiments with the static resource allocation algorithm in SDMA system.

Table B.1: Encoding Order Experiments I

${\sigma_{ij}^3}^2$	**encoding order**	β^1	β^2	β^3	E_C^1	E_C^2	E_C^3	**utility**
0.125	2,1,3	0.5707	0.3156	0.1137	400.25	499.15	720.42	18.7907
0.25	2,3,1	0.6923	0.2005	0.1072	498.54	420.91	686.07	18.7909
0.5	3,2,1	0.7517	0.1882	0.0601	582.77	555.75	524.09	18.9552
0.75	3,2,1	0.7445	0.1893	0.0662	572.25	539.24	673.96	19.1580
1.0	3,2,1	0.7454	0.1991	0.0555	582.39	549.63	786.83	19.3492

Table B.1 displays numerical results obtained with the static resource allocation algorithm for a 3 user MIMO-link-scenario with $d = 3$, where the transmit power budget of the base station is $10W$, the power of the noise is $1W$, QoS-exponents for all links are kept constant at 0.5 (i.e. $\theta^1 = \theta^2 = \theta^3 = 0.5$), the variance of the MIMO channel matrix entries for $l = 1, 2$ are kept at ${\sigma_{ij}^1}^2 = 0.25$ and ${\sigma_{ij}^2}^2 = 0.5$,

and ${\sigma_{ij}^3}^2$ is varied. The wireless channel bandwidth is set to 1 kHz. In this scenario, where QoS diversity is taken out of the question, we clearly see that the algorithm benefits encoding links with better conditions earlier. So, the conclusion is that the better condition (i.e. the higher channel gain) the link has, the earlier it should be encoded.

Similarly, we created another scenario where the variance of the MIMO channel matrix entries are kept constant at 0.5 (i.e. ${\sigma_{ij}^1}^2 = {\sigma_{ij}^2}^2 = {\sigma_{ij}^3}^2 = 0.5$), the QoS-exponents for $l = 1, 2$ are kept at $\theta^1 = 0.25$ and $\theta^2 = 0.375$, and θ^3 is varied (Table B.2).

Table B.2: Encoding Order Experiments II

θ^3	**encoding order**	β^1	β^2	β^3	E_C^1	E_C^2	E_C^3	**utility**
0.125	1,3,2	0.9098	0.0349	0.0553	1068.93	940.54	708.77	20.3876
0.25	1,3,2	0.8893	0.0678	0.0429	972.23	642.83	649.86	19.8264
0.5	3,2,1	0.2211	0.6710	0.1079	473.27	822.49	735.28	19.4769
0.75	3,2,1	0.2154	0.6540	0.1307	471.48	745.77	735.79	19.3760
1.0	3,2,1	0.2129	0.6444	0.1427	470.71	710.93	735.41	19.3258

Clearly, commenting on the results given in Table B.2 is not as easy as it is the case for Table B.1. For $\theta^3 = 0.125$ and $\theta^3 = 0.25$, it is seen that the links with better conditions (i.e. the smaller QoS-exponent or looser QoS requirement) are encoded earlier. However, as θ^3 is increased further, the reverse of the statement becomes true.

So, it is apparently hard to form a general statement on the encoding order decision based on the results of these two sets of experiments. It gets even harder when one set of parameters (either the QoS-exponents or the variance of the channel matrix entries) is not kept at the same value, which is the case for the results displayed on both Table B.1 and B.2.

In Chapter 4, it is shown that the effective capacity for each single MIMO link l can be expressed given its θ^l and ${\sigma_{ij}^l}^2$, i.e. (4.14). Computing the effective capacities of the MIMO links using the MIMO channel rate expression in (4.6) without including interference term, i.e. $R^l = \sum_{i=1}^{d} \log_2 \left[1 + \frac{d^{-1} P \lambda_i^l}{\sigma_n^2}\right]$ using (4.14), they are turned into decision metrics, and inspired by the dominant conclusion for the above pre-

sented experiments, the practical power allocation makes the decision for encoding order of the links as follows:

"The higher the effective capacity a MIMO link has under full power budget and no interference assumption from neighboring MIMO links, the earlier it should be encoded by the base station."

Having a pre-determined encoding order takes the decision variable $\boldsymbol{\psi}$ used in static power allocation algorithm out of the optimization problem definition, which decreases the total number of calculations by a factor of $L!$. We utilize this decision of encoding order in the practical algorithm for SDMA system.

Bibliography

[1] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Trans. Telecomm.*, vol. 10, no. 6, pp. 585-596, Nov. 1999.

[2] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs Technical Journal*, vol. 1, no. 2, pp. 41-59, Nov. 1999.

[3] A. J. Goldsmith, S. A. Jafar, N. Jindal, and S. Vishwanath, "Capacity limits of MIMO channels," *IEEE Journal on Selected Areas in Commununications*, vol. 21, pp. 684-702, 2003.

[4] A. Miu, J. Apostolopoulos, W. Tan, and M. Trott, "Low-latency wireless video over 802.11 networks using path diversity," *In Proceedings of IEEE International Conference on Multimedia and Expo (ICME)*, Baltimore, MD, July 2003.

[5] An Chan, H. Lundgren, T. Salonidis, "Video-aware rate adaptation for MIMO WLANs," *19th IEEE International Conference on Network Protocols (ICNP)*, pp. 321-330, 2011.

[6] D. Wu, and R. Negi, "Effective capacity: a wireless link model for support of quality of service," *IEEE Transactions on Wireless Communications*, vol. 2, no. 4, pp. 630-643, July 2000.

[7] C. H. Liu, A. Gkelias, K. K. Leung, "A cross-layer framework of QoS routing and distributed scheduling for mesh networks," *IEEE Vehicular Technology Conference*, pp. 2193-2197, 11-14 May 2008.

[8] M. Zorzi, J. Zeidler, A. Anderson, B. Rao, J. Proakis, A. L. Swindlehurst, M. Jensen, S. Krishnamurthy, "Cross-layer issues in MAC protocol design for MIMO ad hoc networks," *IEEE Transactions on Wireless Communications*, vol. 13, no. 4, pp. 62-76, Aug. 2006.

[9] J. Liu, Y. Shi, Y. T. Hou, "A tractable and accurate cross-layer model for multi-hop MIMO networks," *In Proceedings of IEEE Infocom 2010*, pp. 1-9, 2010.

[10] K. Mahmood, M. Vehkapera, Y. Jiang, "Delay constrained throughput analysis of a correlated MIMO wireless channel," *IEEE ICCCN 2011*, pp. 1-7, 2011.

[11] K. Sundaresan, R. Sivakumar, M. A. Ingram, T. Y. Chang, "Medium access control in ad hoc networks with MIMO links: optimization considerations and algorithms," *IEEE Transactions on Mobile Computing*, vol. 3, no. 4, pp. 350-365, Oct.-Dec. 2004.

[12] K. Sundaresan, R. Sivakumar, "A unified MAC layer framework for ad-hoc networks with smart antennas," *IEEE/ACM Transactions on Networking*, vol. 15, no. 3, pp. 546-559, June 2007.

[13] K. Sundaresan, R. Sivakumar, "Routing in ad-hoc networks with MIMO links: Optimization considerations and protocols," *Computer Networks*, vol 52, no. 14, pp. 2623-2644, October 2008.

[14] G. Arslan, M. F. Demirkol, Y. Song, "Equilibrium efficiency improvement in MIMO interference systems a decentralized stream control approach," *IEEE Transactions on Wireless Communications*, vol. 6, no. 8, pp. 2984-2993, 2007.

[15] B. Hamdaoui, P. Ramanathan, "A cross-layer admission control framework for wireless ad-hoc networks using multiple antennas," *IEEE Transactions on Wireless Communications*, vol. 6, no. 11, pp. 4014-4024, 2007.

[16] T. Elbatt, "Towards scheduling MIMO links in interference-limited wireless ad hoc networks," *IEEE Milcom 2007*, pp. 1-7, 2007.

[17] B. Mumey, J. Tang, T. Hahn, "Joint stream control and scheduling in multi-hop wireless networks with MIMO links," *IEEE International Conference on Communications*, pp. 2921-2925, 2008.

[18] Y. H. Lin, T. Javidi, R. L. Cruz, L. B. Milstein, "Distributed link scheduling power control and routing for multi-hop wireless MIMO networks," *IEEE Fortieth Asilomar Conference on Signals, Systems and Computers*, pp. 122-126, 2006.

[19] M. O. Pun, W. Ge, D. Zheng, J. Zhang, H. V. Poor, "Distributed opportunistic scheduling for MIMO ad-hoc networks," *IEEE International Conference on Communications*, pp. 3689-3693, 2008.

[20] M. Zhao, M. Ma, Y. Yang, "Opportunistic medium access control in MIMO wireless mesh networks," *Managing Traffic Performance in Converged Networks*, vol. 4516, pp. 458-470, Springer-Verlag, Berlin, 2007.

[21] S. Chu, X. Wang, "Opportunistic and cooperative spatial multiplexing in MIMO ad hoc networks," *IEEE/ACM Transactions on Networking*, vol. 18, no. 5, pp. 1610-1623, Oct. 2010.

[22] X. Liu, N. B. Shroff, E. K. P. Chong, "Opportunistic scheduling: An illustration of cross-layer design," *Telecommunications Review*, 2004.

[23] H. Yin, H. Liu, "Performance of space-division multiple-access (SDMA) with scheduling," *IEEE Transactions on Wireless Communications*, vol. 1, no. 4, pp. 611-618, Oct. 2002.

[24] Y. Yi, M. Chiang, "Stochastic network utility maximization: A tribute to Kelly's paper published in this journal a decade ago," *European Trans. Telecomm.*, vol. 19, no. 4, pp. 421-442, Jun. 2008.

[25] X. Lin, N. B. Shroff, and R. Srikant, "A tutorial on cross-layer optimization in wireless networks," *European Trans. Telecomm.*, vol. 24, no. 8, pp. 1452-1463, Jun. 2006.

[26] L. Georgiadis, M. J. Neely, L. Tassiulas, "Resource allocation and cross-layer control in wireless networks," *Found. Trends Network.*, vol. 1, no. 1, pp. 1-144, 2006.

[27] S. Shakkottai, A. Stolyar, "Scheduling for multiple flows sharing a time varying channel: The exponential rule," *Anal. Methods Appl. Probabil.*, vol. 207, pp. 185-202, 2003.

[28] B. Sadiq, S. Baek, G. de Veciana, "Delay-optimal opportunistic scheduling and approximations: The log rule," *In Proceedings of IEEE Conf. Comput. Commun.*, Apr. 2009, pp. 1692-1700.

[29] L. Tassiulas, A. Ephremides, "Dynamic seRver allocation to parallel queues with randomly varying connectivity," *IEEE Transactions on Information Theory*, vol. 39, no. 2, pp. 466-478, Mar. 1993.

[30] M. Andrews, K. Kumaran, K. Ramanan, A. Stolyar, P. Whiting, "Providing quality of service over a shared wireless link," *IEEE Commun. Mag.*, vol. 39, no. 2, pp. 150-154, Feb. 2001.

[31] M. Andrews, K. Kumaran, K. Ramanan, A. Stolyar, P. Whiting, "Power allocation and routing in multibeam satellites with time-varying channels," *IEEE/ACM Trans. Network.*, vol. 11, no. 1, pp. 138-152, Feb. 2001.

[32] E. M. Yeh, "Multiaccess and fading in communication networks," Ph.D. dissertation, Dept. Elect. Eng. Comput. Sci., Mass Inst. Technol., Cambridge, 2001.

[33] E. M. Yeh, A. S. Cohen, "Throughput and delay optimal resource allocation in multiaccess fading channels," *In Proceedings of IEEE International Symposium on Information Theory*, Jul. 2003, pp. 245-255.

[34] Y. Cui, V. K. N. Lau, R. Wang, H. Huang, S. Zhang, "A survey on delay-aware resource control for wireless systems - large deviation theory, stochastic lyapunov drift, and distributed stochastic learning," *IEEE Transactions on Information Theory*, vol. 58, no. 3, pp. 1677-1701, March 2012.

[35] D. S. W. Hui, V. K. N. Lau, H. L. Wong, "Cross-layer design for OFDMA wireless systems with heterogeneous delay requirements," *IEEE Transactions on Wireless Communications*, vol. 6, no. 8, pp. 2872-2880, Aug. 2007.

[36] J. Tang, X. Zhang, "Cross-layer resource allocation over wireless relay networks for quality of service provisioning," *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 4, pp. 2318-2328, May 2007.

[37] J. Tang, X. Zhang, "Quality-of-service driven power and rate adaptation over wireless links," *IEEE Transactions on Wireless Communications*, vol. 6, no. 8, pp. 3058-3068, Aug. 2007.

[38] J. Tang, X. Zhang, "Cross-layer-model based adaptive resource allocation for statistical qos guarantees in mobile wireless networks," *IEEE Transactions on Wireless Communications*, vol. 7, no. 6, pp. 2318-2328, June 2008.

[39] M. Tao, Y. C. Liang, F. Zhang, "Resource allocation for delay differentiated traffic in multiuser OFDM systems," *IEEE Transactions on Wireless Communications*, vol. 7, no. 6, pp. 2190-2201, June 2008.

[40] R. Berry, "Power and delay optimal transmission scheduling: Small delay asymptotics," *In Proceedings of IEEE International Symposium on Information Theory*, pp. 426, Yokohama, Japan, 29 Jun.-4 Jul. 2003.

[41] R. Berry, R. Gallager, "Communication over fading channels with delay constarints," *IEEE Transactions on Information Theory*, vol. 48, no. 5, pp. 1135-1149, May 2002.

[42] L. Liu, J. F. Chamberland, "On the effective capacities of multiantenna Gaussian channels," *In Proceedings of IEEE ISIT*, 2008.

[43] Q. Du, X. Zhang, "QoS-aware base-station selections for distributed MIMO links in broadband wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 29, no. 6, pp. 1123-1138, June 2011.

[44] C. -S. Chang, "Stability, queue length, and delay of deterministic and stochastic queueing networks," *IEEE Transactions on Automatic Control*, vol. 39, no. 5, pp. 913-931, May 1994.

[45] C. -S. Chang, J. A. Thomas, "Effective bandwidth in high-speed digital networks," *IEEE Journal on Selected Areas in Communications*, vol. 13, no. 6, pp. 1091-1100, Aug. 1995.

[46] J. G. Kim, M. Krunz, "Bandwidth allocation in wireless networks with guaranteed packet-loss performance," *IEEE/ACM Transactions on Networking*, vol. 8, no. 3, pp. 337-349, June 2000.

[47] G. Kesidis, J. Walrand, C. -S. Chang, "Effective bandwidths for multiclass markov fluids and other atm sources," *IEEE/ACM Transactions on Networking*, vol. 1, no. 4, pp. 424-428, Aug. 1993.

[48] A. Schwartz, A. Weiss, *Large Deviations for Performance Analysis: Queues, Communication and Computing*. Chapman&Hall, London, 1995.

[49] A. Goldsmith, *Wireless Communications*. Cambridge University Press, 2005.

[50] A. Bennatan, D. Burshtein, G. Caire, S. Shamai, "Superposition coding for side-information channels," *IEEE Transactions on Information Theory*, vol. 52, no. 5, pp. 1872-1889, 2006.

[51] P. P. Bergmans, T. M. Cover, "Cooperative broadcasting," *IEEE Transactions on Information Theory*, pp. 317-324, May 1974.

[52] D. Tse, P. Viswanath, *Fundamentals of Wireless Communications*. Cambridge University Press, 2005.

[53] L. Massoulie, J. Roberts, "Bandwidth sharing: objectives and algorithms‘," *In Proceedings of IEEE Infocom 1999*, pp. 1395-1403, 1999.

[54] J. Liu, Y. Shi, Y. T. Hou, "A tractable and accurate cross-layer model for multi-hop MIMO ad-hoc networks," *In Proceedings of IEEE Infocom 2010*, pp. 1-9, March 2010.

[55] C. -S. Chang, *Performance Guarantees in Communication Networks*. Springer-Verlag, Berlin, 2000.

[56] M. Fidler, "A network calculus approach to probabilistic quality of service analysis of fading channels," *In Proceedings of IEEE Globecom*, Nov. 2006.

[57] M. Fidler, "An end-to-end probabilistic newtork calculus with moment generating functions," *In Proceedings of IEEE IWQoS*, pp. 261-270, June 2006.

[58] P. Bender, P. Black, M. Grob, R. Padovani, N. Sindhushayana, A. Viterbi, "CDMA/HDR: A bandwidth efficient high speed wireless data service for nomadic users," *IEEE Communications Magazine*, vol. 38, no. 7, pp. 70-77, Jul. 2000.

[59] B. Soret, M. C. Aguayo-Torres, J. T. Entrambasaguas, "Capacity with explicit delay guarantees for generic sources over correlated rayleigh channel," *IEEE Transactions on Wireless Communications*, vol. 9, no. 6, pp. 1901-1911, June 2010.

[60] A. Edelman, N. R. Rao,"Random matrix theory," in *Acta Numerica (2005)*, Cambridge University Press, 2005. pp. 1-65.

[61] I. Krikidis, J. S. Thompson, "MIMO two-way relay channel with superposition coding and imperfect channel estimation," *Elsevier Journal of Network and Computer Applications*, vol. 35, pp. 510-516, 2012.

[62] E. Biglieri, J. Proakis, S. Shamai, "Fading channels: information theoretic and communication aspects," *IEEE Transactions on Information Theory*, vol. 44, pp. 2619-2692, Oct. 1998.

[63] B. Soret, M. C. Aguayo-Torres, J. T. Entrambasaguas, "Evaluation of log moment generation functions for rayleigh channels," Technical Report, Dept. Ing. Commun., University of Malaga, Malaga, Spain, Apr. 2006.

[64] F. P. Kelly, "Notes on effective bandwidths," in *Stochastic Networks: Theory and Applications* (Editors F.P. Kelly, S. Zachary and I.B. Ziedins) Royal

Statistical Society Lecture Notes Series, 4. Oxford University Press, 1996. pp. 141-168.

[65] O. Ercetin, M. O. Memis. (2011, Dec.) "Comments on "capacity with explicit delay guarantees for generic sources over correlated rayleigh channel"," unpublished [Online]. Available: http://arxiv.org/abs/1112.5152.

[66] http://www.isi.edu/nsnam/ns/

[67] A. Iyer, C. Rosenberg, A. Karnik, "What is the right model for wireless channel interference?," *IEEE Transactions on Wireless Communications*, vol. 8, no. 5, pp. 2662-2671, May 2009.