11-28-2011 12:00 AM

Energy Efficient Reduced Complexity Multi-Service, Multi-Channel Scheduling Techniques

Dan J. Dechene, *The University of Western Ontario*

Supervisor: Abdallah Shami, *The University of Western Ontario*

A thesis submitted in partial fulfillment of the requirements for the Doctor of Philosophy degree in Electrical and Computer Engineering

Energy Efficient Reduced Complexity Multi-Service, Multi-Channel Scheduling Techniques

(Spine title: Energy Efficient Scheduling Techniques)

(Thesis format: Monograph)

by

Dan J. Dechene

Graduate Program
in
Electrical and Computer Engineering
Dept. of Electrical and Computer Engineering

A thesis submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

School of Graduate and Postdoctoral Studies
The University of Western Ontario
London, Ontario, Canada

Certificate of Examination

THE UNIVERSITY OF WESTERN ONTARIO
SCHOOL OF GRADUATE AND POSTDOCTORAL STUDIES
CERTIFICATE OF EXAMINATION

Chief Advisor:

Dr. Abdallah Shami

Advisory Committee:

Examining Board:

Dr. Miriam Capretz

Dr. Xianbin Wang

Dr. Mike Katchabaw

Dr. Xuemin (Sherman) Shen

The thesis by
Dan J. Dechene
entitled:
Energy Efficient Reduced Complexity Multi-Service, Multi-Channel Scheduling Techniques
is accepted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Date:

Chair of Examining Board
Dr. Keith Griffiths

Abstract

The need for energy efficient communications is essential in current and next-generation wireless communications systems. A large component of energy expenditure in mobile devices is in the mobile radio interface. Proper scheduling and resource allocation techniques that exploit instantaneous and long-term average knowledge of the channel, queue state and quality of service parameters can be used to improve the energy efficiency of communication.

This thesis focuses on exploiting queue and channel state information as well as quality of service parameters in order to design energy efficient scheduling techniques. The proposed designs are for multi-stream, multi-channel systems and in general have high computational complexity. The large contributions of this thesis are in both the design of optimal/near-optimal scheduling/resource allocation schemes for these systems as well as proposing complexity reduction methods in their design.

Methods are proposed for both a MIMO downlink system as well as an LTE uplink system. The effect of power efficiency on quality of service parameters is well studied as well as complexity/efficiency comparisons between optimal/near optimal allocations.

Keywords: Beamforming, Cross-Layer, Channel State Information, Energy Efficiency, MIMO, Margin Adaptive, Markov Decision Process, Multiuser, Narrow-beam Scattering, Queue State Information, Quality of Service, Resource Allocation, SC-FDMA, Scheduling, Sparse Channels, SVD, Time-Varying Channels.

Acknowledgements

First and foremost, I would like to thank my dearest wife Jessica. Without her love, support and smile, I would not have given up before I started. I am truly happy you are a part of my life and look forward to entering the next step of it with you. The past year apart from each other has been the toughest and most trying things we could both endure. All I can say is that I am profoundly happy that this time is coming to an end, and that we can be together again.

I would also like to thank my parents Doug and Janice. I would not be here without them. Both have always been very supportive, and instrumental in my continuing personal development. Along with that, both my brothers Jon and Jason have been the best bros I could ask for.

I would also like to graciously thank my thesis supervisor Dr. Abdallah Shami for his invaluable insights, guidance and support while pursuing this work.

I would also like to thank my examiners: Dr. Miriam Capretz; Dr. Xianbin Wang; Dr. Mike Katchabaw; and Dr. Xuemin Shen for taking the time to review and examine my thesis.

I would also like to extend my thanks to all my close colleagues and friends from BCIE I've made over the years: Abdessameud Abdelkader, Abdou Ahmed, Ayman Alghamdi, Xiaofeng "Boris" Bai, Mahadevan Balakrishnan, Elham Dolatabadi, Oscar "Bro" Filio, Siamack Ghadimi, Sahar Haghighi, Jamil Hussain, Chris Kennedy, Jason Liu, Marco Luccini, Khalim Meerja, Kevin Mi, Maysam Mirahmadi, Jay Nadeau, Abdulfattah Noorwali, Jahidur Rahman, Andrew Roberts, Tomasz Rybak, Elena Uchiteleva, and Dan Wallace.

I would also like to all the fantastic staff here who have come and gone for helping to make this a great experience. A special thank you to Sandra Vilovski-Anjoli who it has been a joy to get to know over the last many years.

Outside of academics, I can't get by without thanking those friends I've met along the way, particularly those who have made a dramatic impact in my life. From the

extended SOGS crowd, Rebecca Feldman, Duane Jacques, Jonathan Meyer, Kai Pisters, Brent Sherar, and Fiona Simpson are some of the awesome people I've met, through all the excitement, stress and not to mention shenanigans, you won't be forgotten! To those good friends outside of my campus life, Craig Chevalier, Kevin/Anne Fink, Yann/RaeAnne Gagnon and Mike/Jessica Sikora, it's been good times hasn't it? Looking forward to many more years of excitement.

Last and far from least, I would like to thank my good friend Serguei Primak for both his personal and professional advice over the recent years.

Beyond individuals, there has been a number of other contributing factors that were vital to this thesis. With that said, I would like to acknowledge the financial support I received through the University of Western Ontario, the National Sciences and Engineering Research Council (NSERC), the Ontario Government and Broadband Wizard Inc., all of which was vital in giving me the opportunity to pursue my degree.

I would also acknowledge and pay tribute to the Grad Club and the Oar House, which while without them would have stretched funding further, both were an integral part in this success.

Thank you,

Dan J. Dechene

To my loving wife Jessica

Table of Contents

List of Tables

List of Figures

Acronyms

AFD	*Average Fade Duration*
ARQ	*Automatic Repeat reQuest*
BER	*Bit Error Rate*
BLER	*BLock Error Rate*
BPSK	*Binary Phase Shift Keying*
BTFDPS	*Block Time-Frequency Domain Packet Scheduler*
CDF	*Cumulative Distribution Function*
CDI	*Channel Distribution Information*
CMDP	*Constrained Markov Decision Process*
CQI	*Channel Quality Indicator*
CSI	*Channel State Information*
EDCA	*Enhanced Distributed Channel Access*
FDD	*Frequency Division Duplexing*
FDMA	*Frequency Division Multiple Access*
FDPS	*Frequency Domain Packet Scheduler*
FIFO	*First-In, First-Out*
FSMC	*Finite State Markov Chain*
HARQ	*Hybrid Automatic Repeat reQuest*
HCCA	*Hybrid Coordination Function (HCF) Controlled Channel Access*
IEEE	*Institute of Electrical and Electronics Engineers*
IOP	*Information Outage Probability*
LCR	*Level Crossing Rate*
LP	*Linear Programming*
LTE	*Long Term Evolution*
LUT	*Look Up Table*
MA	*Margin Adaptive*

MAC	*Media/Medium Access Control*
MCS	*Modulation and Coding Scheme*
MDP	*Markov Decision Process*
MIMO	*Multiple-Input, Multiple Output*
MPEG	*Motion Pictures Expert Group*
MS-MC	*Multistream, Multichannel*
NLP	*Non-Linear Programming*
OFDM	*Orthogonal Frequency Division Multiplexing*
OFDMA	*Orthogonal Frequency Division Multiple Access*
PAPR	*Peak to Average Power Ratio*
PDF	*Probability Density Function*
PER	*Packet Error Rate*
PHY	*Physical*
PLG	*Power Level Gain*
PMI	*Precoding Matrix Indicator*
QAM	*Quadrature Amplitude Modulation*
QCI	*QoS Class Identifier*
QoS	*Quality of Service*
QPSK	*Quadrature Phase Shift Keying*
QSI	*Queue State Information*
RA	*Rate Adaptive*
RB	*Resource Block*
RLC	*Radio Link Control*
RV	*Random Variable*
SC-FDMA	*Single Carrier Frequency Division Multiple Access*
SISO	*Single-Input, Single-Output*
SNR	*Signal to Noise Ratio*
SVD	*Singular Value Decomposition*
TB	*Transport Block*
TDD	*Time Division Duplexing*
TDPS	*Time Domain Packet Scheduler*
UE	*User Equipment*

V-BLAST	*Vertical Bell Laboratories lAyered Space-Time*
VoIP	*Voice over Internet Protocol (IP)*

Chapter 1
Introduction

Modern wireless communication systems strive to achieve two very contrasting objectives, namely increasing system performance[1] while well as increasing battery lifetime[2]. In fact, both these goals have been heavily focused on since the adoption of early wireless systems; while the definition of system performance has evolved over time, the goal of improving energy efficiency of a system has remained clear.

In early wireless systems, the tradeoff was transparent: increasing transmission power to improve the signal-to-noise ratio (SNR), in turn improved error performance and would inherently require increased energy expenditure. Nowadays, modern technologies such as orthogonal frequency division multiplexing (OFDM), multiple input, multiple output (MIMO) systems, capacity achieving channel coding and even the evolution of signal processing techniques, all contribute to increased scheduler design flexibility. Each one of this technologies represents a *degree of freedom* in the system. The increase in the number of these *degrees of freedom* for a particular system opaques the once transparent tradeoffs in achieving energy efficient communication. Further, the lifting of network stack layer abstractions in recent years through cross-layer design, and consideration of medium access control (MAC) layer performance in the system, takes this a step further as it necessitates examining the overall system performance and energy efficiency, particularly from a per-packet or per traffic stream level. Moreover, when combined with the task of ensuring other quality of service (QoS) constraints of traffic are maintained, the overall energy efficiency tradeoffs have become highly complex.

---

1. The definition and measures of performance will be defined later in this thesis

2. Increasing energy efficiency

Numerous amounts of recent research has focused on areas such as QoS constrained communication, energy efficient scheduling and multiple channel wireless communication individually. Yet, there remains a gap between the large body of research on these individual topics, and a synergized combination thereof. As the adoption rate of streaming video and voice over Internet protocol (VoIP) continues to increase, adequate QoS must continue to be maintained. Modern wireless systems which allow transmission of multiple streams of data such as MIMO or OFDM have long been focused on improving system throughput performance [1–3]. To-date this has helped ensure the ability of wireless systems to meet QoS demands, however such advantages may not be sufficient when energy resources are heavily constrained such as in battery limited wireless devices.

When battery resources are limited in the presence of QoS constraints, the picture changes dramatically. Simple increases in system throughput may no longer be the most desirable method to meet QoS as it may unnecessarily waste energy resources, particularly when the system is not under full load. Here, knowledge of QoS parameters, queue state information (QSI) and channel state information (CSI) play a vital role in the energy efficient scheduler design. However, exploiting all or some of this information exponentially complicates the design of any resource allocation algorithms due to an explosion of the size of the representative state-space. Combining this, with the potential for multiple channel resource allocation may explain the existence of the design gap described above, as energy efficient optimal allocation of resources in a system with such a large *degrees of freedom* is highly computationally complex.

Nevertheless, there has been a small amount of pioneering work in this area to try and approach this problem [4, 5]. This has been accomplished by making various assumptions or simplifications to the system or simply by focusing on a subset of the overall scheduling problem. In this thesis, we propose several different approaches in this research domain.

1.1 Thesis Outline and Contributions

The remainder of this thesis is divided into eight chapters where the technical contributions of this thesis are contained in Chapters 3 to 7 and focus on both downlink and uplink energy efficient scheduling techniques. Prior to introducing the technical contributions, in Chapter 2 we provide detailed background on the problem domain and summarize some of the recent state-of-the-art advances in the area concentrated on in this thesis.

The technical contributions of this thesis begin in Chapter 3 where we propose a novel design for a multiqueue scheduler for a MIMO SVD system with fixed eigenvalues. The proposed method makes use of Markov decision processes to schedule traffic to meet particular delay and throughput requirements while minimizing the average applied transmission power. The major contributions contained in this chapter are

- Proposing dynamic scheduling policy framework for an arbitrary number of queues and channels that meets hard constraints on average delay and throughput which minimizing the average applied power for a time-invariant channel.
- A novel complexity reduction technique for considering joint queue state information through practical limitations by coupling queues through their action-space.
- Proposing a rate-space reduction search by exploiting the monotonicity of bit error rates in symbol rate.

The contributions of this chapter appear in *IEEE Transactions on Wireless Communications* [6].

In Chapter 4, we concentrate on detailing the characteristics of a geometric sparse MIMO channel for use in a later chapter. In this chapter, both time-invariant and time-varying statistics for eigenvalues are derived. Some of the major contributions contained in this chapter are

- Deriving of exact expressions for probability density function of unordered channel eigenvalues, subchannel SNR and capacity.
- Deriving exact expressions for mean and variance as well as tractable approximations to non-closed form equations and verifying their accuracy.

- Computing the exact expression for the level-crossing rate (LCR) and average fade duration (AFD) of the unordered eigenvalues for a single cluster model, as well as approximations for the multi-cluster scenario and verifying their accuracy.

Several of the contributions in this chapter appear in the proceedings of the *25th Queen's Biennial Symposium on Communications* [7].

In Chapter 5 we combine our work in Chapters 3 and 4 in the design of a scheduler with channel distribution information. In addition, the work in this chapter extends Chapter 3 to both account for the time-varying channel and coded modulation schemes. The major contributions contained in this chapter are

- Extending the work from Chapter 3 to account for time-varying channels by exploiting knowledge about the channel distribution information.
- Proposing a practical rate-space reduction by segmenting the symbol rates into three regions and applying subchannel ordering.
- Studying the tradeoff between partition size and time-average applied power.

Part of this work appears in *IEEE Transactions on Wireless Communications* in both [6] and [8].

Chapter 6 which was partially presented at *2011 IEEE International Conference on Communications* [9] looks at the case where multiple traffic streams are generated by multiple users (one stream per user) and multiple channels are available through SC-FDMA. In this chapter, we focus on complexity reduction through establishment of a novel scheduling horizon which considers the restrictions imposed by synchronous HARQ. Both optimal, and suboptimal methods are presented. The major contributions contained in this chapter are

- Proposing a block time, time-frequency allocation scheme that exploits the periodicity of synchronous HARQ to reduce the number of scheduling decisions in time and optimally allocates resources to minimize overall average applied power.
- Proposing two suboptimal approaches to the optimal allocation and compared their performance in terms of average applied power and computational complexity.

- Proposing a tractable least-squares approximation to the information outage probability derived in [10] to efficiently determined the required applied power for a given transport block size and target block error rate.

In Chapter 7 we design a dynamic scheduler for SC-FDMA uplink which meets delay and throughput constraints while minimizing the average applied transmission power. The purpose of this chapter is to describe how to apply some of the methodologies found in Chapters 3, 5 and 6 to meet QoS requirements in SC-FDMA uplink.

Finally, in Chapter 8 we draw some conclusions on this thesis and propose some further research directions in this area.

Throughout this thesis, the reader will find a large number of the system inputs, outputs and decisions are random in nature. As a result, they require knowledge of probability theory. A brief review of applicable theory is given in Appendix A. Additionally, the application of optimization theory plays a vital role in the design of scheduling policies throughout this thesis. An overview of relevant common optimization frameworks are given in Appendix B.

Chapter 2
Background and Literature Review

In this chapter, an overview on the state-of-the-art research related to this thesis is provided including background on the use of both queue state information and channel state information in the scheduling domain.

2.1 Single Stream, Single Channel Transmission Systems

Consider a simplified single stream (or user) system shown in Figure 2.1. This system consists of a single finite-length first-in, first-out (FIFO) queue (or buffer). Packets enter the queue and are stored until transmitted over the channel. Any packets which arrive to the queue when it is full are dropped. Packets which are transmitted over the channel are subject to errors. The probability of a packet being erroneous depends on a number of factors. Packets that are received by the receiver without error are considered successful, while packets arriving in error are dropped as they cannot be decoded.

Packets arrive into the queue at an average rate of $\lambda(t)$ packets per second of size L bits. The queue holds up to B packets. Packets arriving to a full queue are dropped with long-term average probability P_{drop}. The number of packets in the queue at anytime t is $Q(t)$. Packets are serviced from the queue at a rate of $\mu(t)$ packets per second to be transmitted over the channel. Packets are transmitted at a symbol rate of $k(\mu, t)$ with applied power $P(\mu, t)$ over the channel. The channel transfer function which for the majority of this thesis measured as the signal-to-noise ratio (SNR) is time-varying and is described at time t by $h(t)$. Packets arriving at

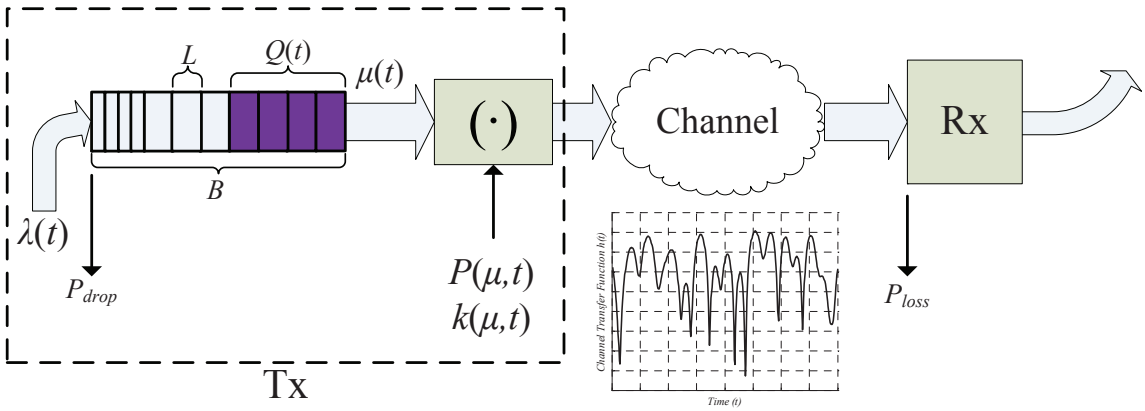


Figure 2.1: Single Queue, Single Channel Transmission System

the receiver arrive successfully with a probability of $1 - P_{loss}$. All of these described quantities can be characterized into three categories namely

1. Fully controllable quantities,
2. Indirectly controllable quantities, and
3. Uncontrollable quantities

where the framework for assigning fully controllable quantities is referred to as a scheduling policy and a given scheduling policy governs the time-evolution dynamics of the above buffer. The symbol $\mathbf{\Omega}$ is used throughout this thesis in reference to a scheduling policy.

Referring back to Figure 2.1, the quantities $\mu(t), k(\mu, t)$ and $P(\mu, t)$ are fully controllable quantities in that they can be adapted or controlled by a given scheduling policy[1]. The quantities $Q(t), P_{drop}$ and P_{loss} are indirectly controllable in that their value depends on the scheduling policy indirectly other quantities (*i.e.,* $Q(t)$ depends on $\mu(t)$ and $\lambda(t)$ and so on). Finally uncontrollable quantities include B and L which are known constants in the system, as well as $\lambda(t)$ and $h(t)$ which are functions of external forces on the system (queue input process and channel physical state respectively).

1. In a saturated transmission scenario where the transmitter is constantly transmitting packets, $\mu(t)$ is directly related to $k(\mu, t)$ through L and the symbol duration which can be denoted T_s.

A scheduling policy defines the manner in which fully controllable parameters are selected or adjusted. This policy can be determined blindly or by exploiting information in the system. Information about $Q(t)$ (queue occupancy level) is known as queue state information (QSI). QSI for example can be used to obtain an estimate of the time packets might wait between being stored in the queue and transmission (herein known as delay). When $Q(t)$ is small, the system has some flexibility to reduce the service rate $\mu(t)$ if it can tolerate some additional delay, while in contrast if $Q(t)$ is large, an increase in $\mu(t)$ might be necessary to meet any constraints on delay. Information about the channel $h(t)$ is also helpful for the scheduling policy and is known as channel state information (CSI). Referring again back to Figure 2.1, it can be observed that the channel fluctuates dramatically as a function of time. Knowledge about CSI for example may allow the scheduler to adjust power to ensure a given packet success rate is achieved, or pause and resume transmission (adjust $\mu(t)$) based on the quality of the channel. In both cases, these advantages are available through knowledge of CSI in the scheduler and are not possible if the transmitter has no CSI.

2.2 Quality of Service (QoS) Constrained Communication

The constraint on delay described above is known as a Quality of Service (QoS) constraint. Constraints on QoS in general can be quantified in a number of ways. Real-time applications such as video, VoIP, telesurgery, and more have necessitated the need to employ network level QoS guarantees on a number of stream parameters. Specific QoS constraints depend widely on the application and can be as simple as defining the minimum throughput required for a particular service, or more comprehensive requirements involving stringent bandwidth and delay tolerances. QoS guarantees in most recent literature fall within constraining one or more of the following quantities

- Delay,
- Jitter,

- Throughput, and/or
- Tolerable packet loss

Further, a particular traffic class may have constraints on the minimum, average and/or maximum of one or all of the above parameters.

2.3 Multi-User/Multi-Stream Scheduling

While it is important to understand the dynamics of the single stream, single channel system, more recent technology usually requires the transmission of multiple streams of data often over multiple non-identical channels.

For example, the accommodation for scheduling multiple streams (or services) for delivery has been incorporated into both the IEEE 802.11e [11] and 802.16e [12]. In addition to these standards, there has been a variety of approaches described [13–15] for multi-stream service delivery. Even within standards, there are several different approaches to handle multistream delivery. For example in the 802.11 family, the 802.11e enhanced distributed channel access (EDCA) focuses on *proportional* service differentiation by maintaining different access parameters for each traffic class. For this reason, 802.11e EDCA is unable to meet hard QoS constraints. While service differentiation is important, it is not sufficient when hard deadline constraints must be met. The centralized hybrid coordination function (HCF) controlled channel access (HCCA) scheme in the 802.11e alternatively provides a framework for allowing wireless stations to meet for hard QoS constraints. The protocol in general however only details the constraint and transmission procedure, but does not describe any channel allocation schemes. More specifically, in the 802.11e HCCA, admission control is used heavily to ensure QoS performance is manageable within a particular set of QoS streams. While the majority of this thesis focuses on scheduling policies rather than admission control, we acknowledge the importance of admission control in achieving QoS in any communication system.

2.4 Multi-Channel Scheduling

Multi-channel communication arises from the recent use of OFDM, MIMO or any technology that enables a system to transmit multiple data streams simultaneously. With OFDM, individual subcarriers of an OFDM symbol may be allocated to different streams of data which in turn may be transmitted to different receivers. In this type of system, the number of available channels is equal to the number of subcarriers.

Multiple channel transmission strategies thatcan be employed in MIMO can be achieved in a number of methods. For example, blind spatial multiplexing techniques (such as V-BLAST [16]) can be employed when state information is not available at the transmitter while eigenvalue-beamforming for example can be used when CSI is available [17]. With some techniques, the maximum theoretical number of sub-channels is equal to the minimum number of antennas between the transmitter and the receiver. In the case of multiple single antenna users, the maximum number of channels is the minimum number of transmitter antennas and single-antenna users.

Irrespective of the method in which multiple channels arise (*i.e.,* OFDM, MIMO spatial multiplexing/eigenbeamforming), individual channels may not be equal and may evolve in time differently from each other. For this reason, multi-channel communication is advantageous from a scheduling perspective as it introduces additional dimensionality to the problem by exploiting these differences in CSI between channels. Here, it becomes possible to select an individual channel (or channels) for transmission of particular packets. This selection may utilize both the CSI vector (CSI for all channels) combined with QoS requirements to select the best channel(s) for transmission. The drawback of multi-channel systems however, is the increased complexity via the introduction of the additional scheduling dimension inherent in having such flexibility.

2.5 Cross-Layer Scheduling Techniques

The concept behind cross-layer scheduling techniques is rather simple: The more information one has, the better decision one can potentially make. This is rather

intuitive. Similar to the case of the single stream, single channel case, multi-stream, multi-channel scheduling can benefit from information about the channels and queue occupancy levels are of importance in scheduler design.

2.5.1 Channel State Information (CSI)

It has been known for some time [18] that channel state information (CSI) which is knowledge about the current channel condition, can improve scheduling performance.

A simple example to see an improvement, is the ability of MIMO to offer increased channel capacity when full CSI is known. The capacity of a MIMO channel is given in [17]

$$C = \sum_{i=1}^{M} \mathbb{E}\left[\log\left(1 + \frac{P(\lambda_i)}{N_0}\lambda_i^2\right)\right] \tag{2.1}$$

where $M = \min(M_R, M_T)$, λ_i^2 is the i^{th} eigenvalue of the MIMO channel matrix, P is the total system transmission power, $P(\lambda_i)$ is the power allocated to the i^{th} channel with M_R and M_T denoting the number of receiver and transmitter antennas respectively. The benefits of CSI are observed in adoption of a power allocation policy for $P(\lambda_i)$ by the transmitter. If a transmitter has no information about CSI (*i.e.,* the value of these eigenvalues), the best power allocation strategy is to simply allocate power equally across all channels (*i.e.,* $P(\lambda_i) = P/M_T$). Alternatively, when the transmitter has full CSI, the transmitter has the ability to allocate power based on these channel eigenvalues. It has been shown [17] that the optimal power allocation in this case is the adaptive waterfilling approach resulting in a channel dependent power allocation of

$$P^*(\lambda_i) = \left(\mu - \frac{N_0}{\lambda_i^2}\right)^+ \tag{2.2}$$

where $(x)^+ = \max(0, x)$ and μ is the adaptive waterfilling level satisfying

$$\sum_{i=1}^{M} \mathbb{E}\left[\left(\mu - \frac{N_0}{\lambda_i^2}\right)^+\right] = P \tag{2.3}$$

While the improvements afforded through adaptive waterfilling depend largely on the eigenvalue distribution, noise power and total power allocated, the system offers increased capacity by exploiting the ability to choose a power allocation scheme that depends on the individual eigenvalues and how they change in time rather than simply allocating power blindly.

As shown there is a large benefit to exploiting full CSI at the transmitter, however in general it is complicated to ensure that this information is available at the transmitter error-free, for all time. In practice, CSI is only available at the transmitter due to feedback from the receiver (either through a direct feedback channel or through assuming reciprocity of the channel) and in general this feedback could be delayed. Further, transmission of the feedback for all time may require a large amount of transmission bandwidth.

Finite-state Markov channel (FSMC) modelling [19] has been used heavily exploited as a tool in scheduler design. FSMC modelling reduces the feedback requirements. Rather than considering the channel as a continuous quantity, where in the above, a continuously adaptable MIMO power allocation policy is employed and large amounts of information is needed as feedback, with a FSMC model the receiver only relays information about which state (or region) the channel falls under, and the transmitter chooses a suitable transmission policy corresponding to that state. While the overall performance gain is limited when compared to full, continuous CSI, the amount of feedback and decision complexity is generally much lower. An overall of FSMC modelling is given in Section A.4 of Appendix B.

2.5.2 Channel Distribution Information (CDI)

It is possible that information about the current channel state (CSI) is unavailable, for example due to implementation limitations. While this may limit adaptability of the scheduler when compared to CSI, channel distribution information (CDI) can still be exploited in scheduler design for some performance improvements. CDI refers to information about the overall statistic properties of the channel (*i.e.,* distribution of

the channel SNR). This information could be obtained in a number of ways a priori such as through measurement campaigns or use of statistic channel models.

2.5.3 Queue State Information (QSI)

The exploitation of queue state information (QSI) has been shown to be beneficial particularly when transmitting QoS constrained traffic and when combined with exploitation of CSI [20, 21].

Consider the following scenario when the channel is in a bad condition (in a state when it will take a large amount of transmission power to transmit with a certain probability of error). Without knowledge of QSI, the best policy is to simply postpone transmission until the channel is in a good state to conserve transmission power. Now suppose upon analysis of the queue (utilizing QSI), the scheduler determines that it is not possible to meet QoS constraints if transmission is delayed all together. In this case, the transmitter may choose to transmit anyways at an increased power level to ensure QoS constraints are met.

Unlike CSI, QSI for finite length queues can be completely described by a finite number of states as at any instant of time, only a certain number of packets are stored in the queue (unlike the continuous nature of CSI). When a system contains multiple queues, the system QSI is completely described by the joint occupancy levels in all queues. The resulting state-space grows exponentially as $\mathcal{O}(B^K)$ where B is the number of states a queue can exist in and K is the number of queues.

As with CSI, it is possible to quantize QSI (or only consider partial QSI). For example, in work by Berry and Gallager [22], later extended by Neely [23] to multiple queues, queue(s) can be segmented in two states (left and right of a threshold). In this case, a scheduler can employ two policies based on the current state of the queue relative to the threshold. Both policies strive to return to the threshold queue level. While the number of policy actions taken by a scheduler is limited in this case, the complexity growth is only $\mathcal{O}(2^K)$ providing a more tractable policy complexity.

2.6 Multi-Channel, Multi-Stream Optimization

Multi-channel, multi-stream optimization arises in numerous practical scenarios in wireless communications such as QoS scheduling and spatial multiplexing problems.

For details on design considerations, consider a very general downlink system as shown in Figure 2.2 consisting of a base station and K users (or a single user with K traffic streams). Here there are K queues (or K users each with a single queues) and L parallel transmission channels. In this downlink system, all transmissions originate from a single base station to a single user (or K independent users) over one or more parallel subchannels. The full details of this model are described as follows.

Each of the K traffic streams arrives to an independent FIFO queue that can store up to B_i packets. Each traffic stream has a set of quality of service (QoS) parameters associated with it $\{D_i, \bar{D}_i, \lambda_i, \delta_i, J_i, \bar{J}_i, L_i\}$. Here D_i and $\bar{D}_i$ denote the maximum tolerable instantaneous and average queueing delay respectively. The quantities J_i and $\bar{J}_i$ denote the maximum tolerable instantaneous and average single-hop jitter respectively and δ_i denotes the average tolerable loss rate in packets for a traffic class. Finally, λ_i is the average arrival rate in packets to the i^{th} queue and L_i is the packet length. Any packet arriving to a full queue is dropped. The arriving packets follow a general arrival process.

Each of the K queues have a time-dependant queue service rate of $\mu_i(t)$ packets per second. For all time t, there are L channels available for transmission. Let $C_{i,l}(\boldsymbol{\mu}, t)$ denote the rate of packets taken from queue i and transmitted over channel l and $\mathbf{C}_l(\boldsymbol{\mu}, t)$ be a vector denoting the combination of packets taken from all queues transmitted over single channel l. Here, $\boldsymbol{\mu}$ is the vector of all queue service rates $\mu_i, \forall i$. During any time t, the number of packets stored in any queue is given as $Q_i(t)$ where $Q_i(t) \in \{0, 1, \ldots, B_i\}$.

For each channel, the transmitter can adjust both the power allocated $P_{i,l}(\boldsymbol{\mu}, t)$ and the transmission rate $f_{i,l}(\mathbf{C}_l(\boldsymbol{\mu}, t))$ for a given channel and given traffic stream transmitted in that channel. For simplicity, $\mathbf{P}_l(\boldsymbol{\mu}, t)$ and $\mathbf{f}_l(\mathbf{C}_l(\boldsymbol{\mu}, t))$ denotes these quantities in vector form for all i. Without loss of generality, we assume that power can be allocated continuously in both time and dynamic range such that

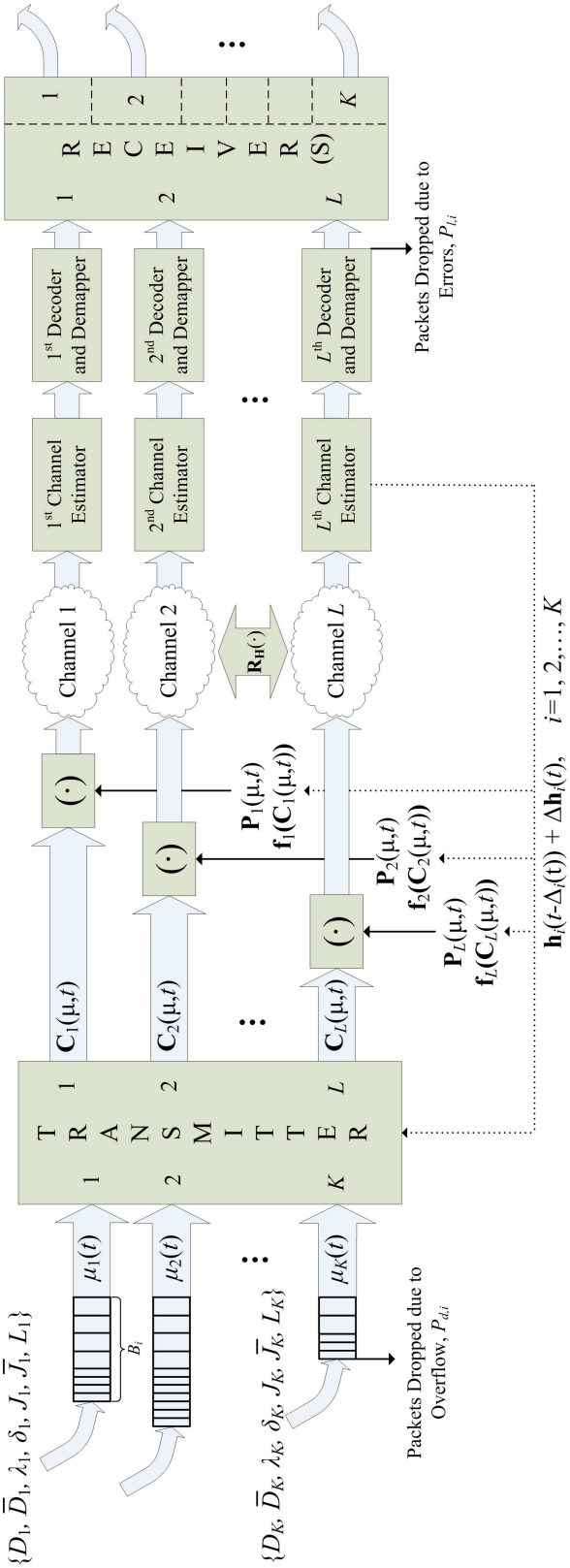


Figure 2.2: General System Model for Multi-Stream, Multi-Channel Transmission System

$0 \leq P_{i,l}(\boldsymbol{\mu}, t) < \infty, \forall i, l$. Further we also assume the rate of the channel, measured in bits/second, can be adjusted continuously such that $0 \leq f_{i,l}(\mathbf{C}_l(\boldsymbol{\mu}, t)) < \infty, \forall i, l$.

The channel estimation of the l^{th} channel for traffic class i is denoted $h_{l,i}(t)$. The set of channel all measurements for a single class i is $\mathbf{h}_i(t)$ and the matrix of all $\mathbf{h}_i(t)$ for all l is $\mathbf{H}(t)$. The transmitter obtains a delayed, imperfect estimate of $\mathbf{h}_i(t)$ given by $\mathbf{h}^{(T)}(t) = \mathbf{h}_i(t - \Delta_i(t)) + \Delta\mathbf{h}_i(t)$ where $\Delta_i(t)$ is the time-dependant delay of the feedback channel and $\Delta\mathbf{h}_i(t)$ is the estimation error. The L channels also may or may not be correlated. All erroneous packets arriving at the receiver(s) are dropped.

2.6.1 Scheduling Policy Considerations

The above describes a very general system downlink wireless communication system. This system from a resource allocation point of view is analogous to many multi-queue, multi-channel systems encountered in the literature such as a multi-user OFDM downlink/uplink system or a single base station and receiver pair employing SVD eigenbeamforming or spatial multiplexing methods such as V-BLAST.

In general the designer must manage resources, such as packet allocation, power assignment and rate adaptation. An overall allocation routine is referred to as a scheduling policy, which throughout this thesis is denoted as $\boldsymbol{\Omega}$. A designer may wish to exploit system information such as channel feedback (CSI), information about the occupancy level of the queues (QSI) and QoS parameters in the design of the scheduling policy in order to achieve a particular type of, and level of performance. Each scheduling policy has specific performance goals such as ensuring QoS constraints are met, providing fairness, maximizing throughput and minimizing energy consumption. In general a designer may choose to focus on some or all of these goals, or as some of these goals are conflicting in nature, target a trade-off between them.

Each of the above design decisions represents a *degree of freedom* in the scheduling policy design. In general it is difficult for a designer to exploit every *degree of freedom* described above as the system complexity increases dramatically for each additional *degree of freedom* introduced in the policy design. This has been referred

to as the *Curse of Dimensionality* [24]. In the literature, there are usually a number of simplifications or tradeoffs that are considered. In the following section, we will present a review of related schemes and details on how they differ in terms of addressing the downlink resource allocation problem described above.

In order to understand the intricacies involved in the design of a scheduling policy, we first present the energy-optimal general optimization formulation of the above downlink scheduling problem. While we note it may not be possible to obtain the energy-optimal solution with any known framework within a feasible degree of complexity, this general model provides a benchmark to facilitate description of the recent literature in this area as well as potential future areas of research.

First, the objective function of the energy-optimal scheduling policy describing the average power consumption of the system is

$$\psi_{avg} = \min_{\substack{f_{i,l}(t), P_{i,l}(t), \\ C_{i,l}(t), \forall i,l,t}} \left(\lim_{t\to\infty} \frac{1}{t} \int_0^t \sum_{i=1}^{K} \sum_{l=1}^{L} \frac{P_{i,l}(t) C_{i,l}(t)}{f_{i,l}(t)} dt \right) \tag{2.4}$$

where all quantities are defined as before. The optimal scheduling policy, $\mathbf{\Omega}^{opt}$, describes $f_{i,l}(t)$, $P_{i,l}(t)$, and $C_{i,l}(t)$ for all i, l, t to minimize (2.4) while not violating QoS constraints. While the solution to the above will necessarily yield an energy-optimal scheduling policy, such a policy is not guaranteed unique nor computationally tractive for unbounded time t. While greedy-based approaches may be applicable in such unbounded time-domain problems, they may not be provably optimal or near optimal for such a general problem described above; particularly with underlying constraints on the allocation quantities.

Here, we make two reasonable assumptions on the expression for ψ_{avg}. Firstly, we note that in most wireless communication systems, particularly packet systems, allocation decisions are made at discrete intervals, which we denote as frames. While it is not necessary to employ a fixed size frame (*i.e.,* the 802.11 distribution coordination function [25] employs no fixed frame for transmission), the transmission decision thresholds in general are discrete in time. With the assumption of relatively small decision frame sizes of duration T_f seconds, there is little loss in generality to assume

that the general system above can employ fixed-size decision thresholds. As $T_f \to 0$, this approaches the continuous time system of (2.4). Using this argument, we can rewrite (2.4) as

$$\psi_{avg} = \min_{\substack{f_{i,l}(n), P_{i,l}(n), \\ C_{i,l}(n), \forall i,l,n}} \left(\lim_{m\to\infty} \frac{1}{mT_f} \sum_{n=0}^{m} \sum_{i=1}^{K} \sum_{l=1}^{L} \frac{P_{i,l}(n) C_{i,l}(n)}{f_{i,l}(n)} \right) \tag{2.5}$$

In (2.5), it is clear that while we established finite decision thresholds, the overall optimization is still unbounded in time. In general, the evolution of any one parameter in time (including the channel evolution) is bounded to a certain degree. As such, it has become common practice to describe the evolution of a system in time as a finite number of states rather than time evolution [20, 26, 27]. Here we make no assumptions on the number or definition of such states, but suggest that in general it is possible to describe the system in this manner. Let $\mathcal{S}$ be the set of all possible states such that $s \in \mathcal{S}$. The following two prepositions are true about $\mathcal{S}$.

1. During anytime m, the system can exist in a single state $s \in \mathcal{S}$, and
2. The set $\mathcal{S}$ is a complete set such that proposition 1 is true for all m

Using the above definition, we modify (2.5) to obtain

$$\psi_{avg} = \min_{\substack{f_{i,l}(s), P_{i,l}(s), \\ C_{i,l}(s), \forall i,l,s}} \left(\sum_{s\in\mathcal{S}} \Pr[S = s] \sum_{i=1}^{K} \sum_{l=1}^{L} \frac{P_{i,l}(s) C_{i,l}(s)}{f_{i,l}(s)} \right) \tag{2.6}$$

where $\Pr[S = s]$ is the probability of the system existing in state s. For simplification, let $p^{(s)} = \Pr[S = s]$[2]. While in general, the above problem remains highly non-trivial to solve, it is no longer a function of time. In the following sections, we briefly discuss the constraints for the above problem.

2. While the revised formulation is no longer unbounded in time, it is important to note that $p^{(s)}$ is not necessarily independent of the allocation quantities

2.6.2 Average Delay

In general [28], one can obtain a simple expression for average delay by noting that the average service rate of a finite length queue is equal to the average arrival rate of packets entering the queue. Using this argument, the delay constraint is simply

$$\bar{D}_i \geq \bar{\mathcal{D}}_i = \frac{\mathbb{E}[Q_i(s)]}{\mathbb{E}[\mu_i(s)]} = \frac{\sum_{s\in\mathcal{S}} p^{(s)} Q_i(s)}{\sum_{s\in\mathcal{S}} p^{(s)} \mu_i(s)}, \quad \forall i \tag{2.7}$$

where $\mu_i(s)$ and $Q_i(s)$ is the service rate of and the number of packets in queue i respectively while in state s and $\mathbb{E}[\cdot]$ is the expectation operator where the expectation is taken over $\mathcal{S}$.

2.6.3 Maximum Delay

In short, it is difficult to design an energy-optimal schedule for a maximum delay bound. The reason for this is two-fold. First, in general it may not be possible to ensure such a bound is met while not considering the time-evolution of each packet (in time, not state). Such a method implies a un-bounded optimization over time is required. There are unbounded methods to approach such a problem [29], however optimization in this case must be handled online. Secondly, due to the general discipline of the arrival rate, and potential limitations on the service rate of each channel, there may exist a non-zero probability of the instantaneous arrival rate over a short period of time exceeding the maximum service rate. As such, it may be infeasible to ensure the maximum delay bound is met for all time particularly for finite transmission power. For this reason, it is often assumed $D_i = \infty, \forall i$.

2.6.4 Average Tolerable Packet Loss

In the above system, packets can only be lost in two ways. First, any packets arriving to a full queue are dropped; this is known as packet dropping. Second, it is also possible that the receiver may not be able to decode packets as a result of bit errors,

this is known as channel losses. The total loss rate comprises both types of losses. Denoting $P_{d,i}(s)$ and $P_{l,i}(s)$ respectively as the average queue drops and channel loss rates in state s, the average total tolerable packet loss rate is given as

$$\begin{aligned} \delta_i &\geq \zeta_i = 1 - (1 - \mathbb{E}[P_{d,i}(s)])(1 - \mathbb{E}[P_{l,i}(s)]), \quad \forall i \\ &\geq 1 - \left(1 - \sum_{s \in \mathcal{S}} p^{(s)} P_{d,i}(s)\right) \left(1 - \sum_{s \in \mathcal{S}} p^{(s)} P_{l,i}(s)\right) \end{aligned} \tag{2.8}$$

2.6.5 Instantaneous Power

Both hardware or regulatory restrictions may place limitations on the total transmit power at any point in time. Denoting the overall limit on power as P on average over a state s, the power is constrained by

$$\sum_{i=1}^{K} \sum_{l=1}^{L} \frac{P_{i,l}(s) C_{i,l}(s)}{f_{i,l}(s)} \leq P, \forall s \in \mathcal{S} \tag{2.9}$$

2.6.6 Throughput

The throughput constraint in general is related directly through the traffic stream arrival rate and the tolerable loss rate. The throughput constraint is simply

$$\chi_i \geq \lambda_i (1 - \delta_i) \quad \text{packets per second} \tag{2.10}$$

Due to the relation of system throughput to the tolerable loss rate and the arrival rate, it is not necessary to constrain both the throughput and the tolerable loss rate, rather constraining one is sufficient.

2.6.7 Jitter

In a general traffic model, individual streams may have limits on the maximum tolerable jitter (J_i) and the average stream jitter ($\bar{J}_i$). Jitter can negatively affect the performance of real-time traffic as real-time decoders are often designed to receive data at a constant rate such as MPEG2-TS [30].

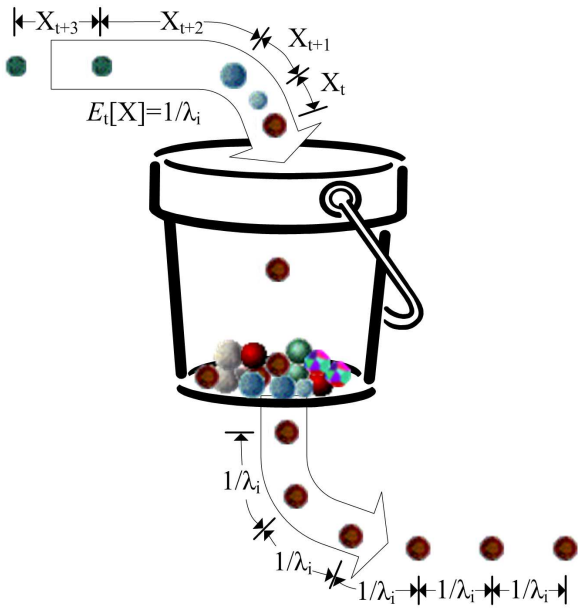


Figure 2.3: Leaky Bucket Jitter Compensation

While scheduling algorithms have a large impact on the single-hop jitter, it is difficult in general to design a system to accommodate this directly. Most applications employ a leaky bucket based system at the receiver to mold incoming traffic into the desired bit-rate and combat jitter. The conceptual diagram of a leaky bucket system is shown in Figure 2.3.

Here we see that by introducing some additional delay in the system in the form of storing packets in a bucket in such a way that the bucket is never empty, one can essentially remove jitter in a system. Therefore, in the general formulation we neglect jitter by assuming $\bar{J}_i = \infty$ and $J_i = \infty, \forall i$.

2.7 Recent Relevant Research Work

In order to assess the current state of the art, we review recent approaches to the above energy-optimal problem. In general, there are few works found in the literature that attempt to address the above problem directly. This is due to the complexity described in the above problem when considering both QSI and CSI. There are however a number of works that focus on addressing subsets of the entire problem. We

review these and relevant research here. The major portion of this review focuses on packet based systems.

The remainder of this section is divided as follows. In Subsection 2.7.1, we overview existing research that contributes to improving energy-efficiency in the multi-stream, multi-channel (MS-MC) scenario indirectly, while in Subsection 2.7.2 we overview recent advances to solve the above described problem specifically from a scheduling perspective and draw comparisons to the general system model previously presented.

2.7.1 Throughput Efficiency Based

There is a large amount of research work focused on achieving throughput efficiency, particularly for MS-MC systems. More specifically, the sum-rate of a system is maximized in the majority of these schemes. The sum rate is the total achievable transmission rate of a system and is contributed by transmissions across all channels and/or to all users.

Application of sum-rate maximization schemes is particularly important for energy-efficiency communication in saturated wireless systems[3]. In these scenarios, maximizing the sum-rate for a particular power level is of importance when examining the transmission energy required per bit of information, as increasing the transmission rate for fixed power increases the system energy efficiency.

The majority of these schemes can be classified into

1. Precoder (code) design
2. User selection (resource/channel selection), and
3. Power/rate control

While these schemes are not described in detail here as they do not directly relate to the work in this thesis, Table 2.1 highlights some of the recent pioneering works that focus on sum-rate maximization for the MS-MC downlink for the interested reader.

3. Here we denote saturated wireless systems as systems that necessitate transmission at the maximum transmission rate at all times.

Table 2.1: Comparison of Pioneering Sum-Rate Maximization Based Multi-Stream/User Resource Schemes

Ref.	Classification	Details
[31]	*Survey*	Overview of sum-rate maximization resource assignment schemes based one user scheduling, precoding and power-allocation.
[3, 32]	Precoding	Methods of sum-rate maximization through employment of user ordering and successive user stream encoding based on knowledge of interference to each user in a Gaussian broadcast channel by employing dirty paper (DP) coding [33].
[34]	Precoding	Tomlinson-Harashima non-linear precoding scheme is employed (extension some single user/antenna design) and achieves higher sum-rate than linear or decision feedback equalizers.
[2]	User Selection	Dynamic subcarrier allocation for OFDM to maximize sum-rate
[35–37]	User Selection	Selection of a subset of users to maximize sum-rate when number of users exceeds number of base-station antennas for DP coded systems
[38, 39]	User Selection	Selection of a subset of users to maximize sum-rate when number of users exceeds number of base-station antennas for block diagonalization (BD) systems
[40, 41]	Power and Rate Control	Iterative waterfilling methods for power allocation approach to maximize sum-rate
[42]	Power and Rate Control	Proposed power and rate allocation methods to maximize weighted sum-rate and minimize sum power for OFDMA

Application of dynamic backpressure policies [43] is also of interest in throughput maximization for MS-MC downlink systems. Such policies are well-known to be throughput-optimal. In the work by Manikandan et al. [44] for example, the authors employ a backpressure based policy to the MS-MC resource allocation problem. Here, the authors propose an extension to previous works by [45] and [46] and design a method to assign rates based on partially known CSI. During each scheduling time-frame, only one user is assigned to each channel. The author's formulation of error rates is based on channel capacity outages. A packet is successfully received when a given rate is less then the channel capacity, and received in error otherwise. As such, power control is not employed or considered in this work, alternatively it is integrated into determining the achievable capacity region. Finally, the formulation is not constrained by any delay or loss requirements.

2.7.2 Energy Efficiency Based

There have also been a number of recent works that have attempted to simplify and solve the energy-optimal allocation problem described above (or a closely related problem). These schemes are detailed below and a summary of key scheme components is given in Table 2.2.

Rashid et al. [4] proposed a scheduling scheme to exploit the diversity present when using V-BLAST [16] to transmit to multiple users. Here, to simplify the design, the channels $\mathbf{h}_l$ are broken up into a small number of states based on the post-processed SNR of each sequential V-BLAST stream which is also given from [16]. The number of streams is equal to the number of transmitter antennas on the base station. As V-BLAST operates in open loop MIMO, the transmitter is only made aware of each channel state rather than the exact estimate of the channel. Power is allocated equally across all channels ($\mathbf{P}_l(\boldsymbol{\mu}, t) = const$). Further, there is a finite number of transmission rates such that the number of transmission rates is equal to the number of channel states. The transmission rate is chosen solely based on the channel state (*i.e.,* $\mathbf{f}_l(\boldsymbol{\mu}, t) = \mathbf{f}_l(\mathbf{H})$) where each transmission mode is statically mapped to a channel state and where the lowest SNR channel state has a null transmission rate of

0 bits/second. The active transmission rates are chosen in such a way as to satisfy a minimum target packet error rate (PER).

Arrivals to each user follow a batch bernoulli arrival process. During each scheduling interval, only one subchannel is allocated to a given user implying

$$C_{i,l} = \begin{cases} \sum_{i=1}^{K} C_{i,l}, & \text{channel is allocated to user} \\ 0, & \text{otherwise} \end{cases} , \forall l \tag{2.11}$$

The user allocation is handled by exploiting multiuser diversity. During each scheduling interval, each subchannel is allocated to the user with the highest SNR state. When two or more users share the highest SNR state, a user is randomly chosen from this group. Delay and packet loss is not constrained, however performance metrics for the delay and loss metrics are derived.

Niyato et al. [5] approach the problem from a similar direction using MIMO SVD eigenbeamforming to obtain multiple parallel channels. In their work, the entire problem is segmented into separate admission control and resource allocation problems; the latter of which we describe here. Power and rate allocation is handled as described above in [4] where the channel is partitioned based on the MIMO channel eigenvalues. The antenna assignment (or resource allocation) is handled through an exhaustive allocation action space which spans the total number of ways the transmission antennas can be grouped and assigned to all users. As in the scheme above, one channel can only be occupied by a single user during a time frame. While the exhaustive method is efficient for a small number of users, the action space grows considerably as the number of users increases.

The resource allocation is formulated by applying a direct application of Markov decision process (MDP) theory where MAC layer throughput is maximized (through minimizing of queue losses) and average delay is constrained (as in (2.7)). The drawbacks of this approach however, is the complexity issues in consideration of the joint state space of the individual queues and channels.

Lau with Chen [48] and with Cui [47] formulate the optimization problem as a Pareto-optimal tradeoff between delay and power for multi-stream MIMO and

Table 2.2: Summary Multi-Streams, Multi-Channel Resource Allocation Schemes

Ref.	Power Allocation	Rate Adaption	User/Channel Allocation
[4]	Fixed	Fixed to Channel State	Highest User SNR
[5]	Fixed	Fixed to Channel State	Allocation that maximizes throughput through exhaustive search
[47]	Variable	Function of Power/CSI	Based on MDP solution using QSI/CSI
[48]	Variable	Function of Power/CSI	Precoder Design
[29]	Variable	Function of Power/CSI	———

OFDMA systems respectively. The Pareto-optimal formulation in both works is shown to be delay-optimal.

In [47], the authors assume the system operates at channel capacity in all OFDM subcarriers. In this way, for a given CSI $\mathbf{H}$, there is a one-to-one mapping between the transmission rate $\mathbf{f}_l(\mathbf{C}_l(\boldsymbol{\mu}, t))$ and the subcarrier power allocation $\mathbf{P}_l(\boldsymbol{\mu}, t)$. The subcarrier power and allocation problem is formulated as a delay-optimal infinite horizon MDP for the joint QSI and CSI. To reduce the system complexity the authors propose two methods. Firstly, the authors propose a reduced state method to solve the Bellman equations by solving separate conditional Bellman equations. The latter, is a two-step procedure where the subcarriers are allocated subject to CSI only, and power is allocated subject to both CSI and QSI.

In [48], the author incorporate stream error rates by taking an approach similar to [49] to relate the transmission rate directly to the subchannel power allocation through introduction of a constraint in the rate expression. The optimization problem is formulated as in [47] to obtain the optimal precoding matrix. In this way, user selection is not required (all users transmit during each frame, *i.e.,* $\mathbf{C} = \text{diag}\{\mu_1, \ldots, \mu_K\}$). The authors also propose a suboptimal, low complexity extension that exploits the relative delay importance (β factor) of each stream.

The drawback in both papers is the lack of hard constraints on the stream delay. While the delay-optimal method provides differentiation between streams through specification of the β factor, this method does not directly translate into hard delay

constraints.

Kuo and Cavers take an interesting approach to the rarely tackled problem of maximum delay constraints in [29]. In their work, by employing resource and power allocation, they are able to meet maximum delay constraints and minimize transmission energy for the special case of a single traffic stream (*i.e.,* for $K = 1$) over multiple parallel channels. To combat the time-dependant issues on meeting hard delay-constraints, the author's optimization constraints are modified in time domain to force inequality of constraints at the interior points and equality of constraints at the problem upper bound, such an algorithm however is non-causal.

In their work, channel error constraints are met by introducing a constant gain factor in the capacity formulation as computed in [49] to meet a target bit error rate (BER). In this way, as with other schemes we have described above, there is a one-to-one mapping between the power and rate allocation for given CSI. In lieu of exploiting direct QSI, the authors utilize knowledge of the current expiring data (data which will exceed maximum delay requirements if not transmitted) during a given frame.

The authors further propose a causal version of the scheduler based on estimates future energy usage based on current CSI and data expiry. The causal scheduler performs the above optimization at each instant of time a new packet arrives to the system or when the CSI changes. The drawback of course of such an approach is that CSI may change rapidly, and packets may arrive to the system regularly forcing the online optimization procedure to be performed frequently. The scheduling scheme is also limited to the case of a single traffic stream.

2.8 Chapter Summary

While, in recent years, there has been a large number of contributions to the MS-MC problem particularly in improving system throughput, in the realm of constraining quality of service, there has been far less focus. One notable rationale for this is the well-known complexity issues associated with exploitation of QSI, particularly for multistream transmission. Many of the above described schemes rely on underlying

assumptions or system simplifications to achieve complexity reduction. Future approaches will need to continue to focus on methods of achieving complexity reduction without dramatically sacrificing system performance. In general, practical systems place limitations on design flexibility, such limitations could be exploited in scheduler policy design to reduce system complexity. It is this ideology which is exploited throughout this thesis.

Chapter 3
Energy Efficient Downlink Bit-Level Scheduler for Static MIMO Wireless Links

In the previous chapters, we highlighted the importance of cross-layer protocols, particularly in meeting QoS constraints. Recently, there has been a large body of research surfacing which employs the cross-layer ideology to meet QoS constraints in the wireless environment [20, 21, 48, 50–54]. In [51] and [20] for example, the authors focus on constrained QoS where they show that knowledge of the instantaneous buffer occupancy, also known as queue state information (QSI), combined with knowledge of the wireless channel state can minimize the average transmission power subject to constraints on MAC layer throughput and average delay. These works however are mainly focused on a single queue which is serviced over a single channel which we herein denote these as $\{1 \times 1\}$ systems. Most modern wireless communications systems however, are comprised of multiple input queues with various QoS requirements, and can be transmit over multiple parallel channel (such as those channels provided by multiple antenna or MIMO systems). More generally, we define a $\{K \times M\}$ system as a transmission system with K queues as inputs to the system and M is the number of parallel channels available for transmission (as discussed in Chapter 2. While several recent parallel works [47, 48, 50, 54] have looked at scheduling techniques for supporting QoS in $\{K \times M\}$ systems, to the best of our knowledge there exists no prior works which exploits full QSI in these systems while meeting hard, heterogeneous QoS constraints. In Lau and Chen's work [48] for example, a framework for a delay-optimal power and resource allocation is proposed for such a heterogeneous traffic system, however the weights employed for delay in the optimization framework do not impose hard guarantees on heterogenous average delay and losses which is required for QoS stringent traffic streams.

In this chapter, we propose a design for a cross-layer scheduler in the presence of full QSI for a generic $\{K \times M\}$ system to target specific average delay and packet loss rates. Through application of a novel MAC layer rate assignment scheme, we design a scheduler that is able to exploit full QSI to meet QoS requirements while reducing long-term average power consumption and reduced complexity compared to the full-scale optimization problem.

The remainder of this chapter is divided into 5 sections. First, in Section 3.1 we detail both the MAC and PHY layers which are used throughout the chapter. The MAC layer consists of an arbitrary number of finite queues with varying delay and throughput requirements. Packets from these queues, are serviced over multiple channels provided by MIMO singular value decomposition (SVD).

Section 3.2 presents the details of the scheduler design. Due to the problem complexity of considering the QSI of all queues simultaneously, this chapter presents a method of decoupling the problem into a two-step scheduling policy by exploiting practical limitations on the system, namely the packet service rate.

Next, in Section 3.3, the details of how to formulate the framework from Section 3.2 is formulated using traditional optimization frameworks. Section 3.4 provides detailed simulation results on the average power allocation as a function of key quantities and finally Section 3.5 summarizes the key findings of this chapter. These key findings of this chapter are also found in Dechene et al. [6].

3.1 System Model

The system model used in the work is a general $\{K \times M\}$ downlink model where K is used to denote the number of independent MAC layer queues as an input to the system and M is used to denote the number of PHY layer channels available for transmission. The overall system model is shown in Figure 3.1 where the MAC and PHY layer subcomponents are discussed below.

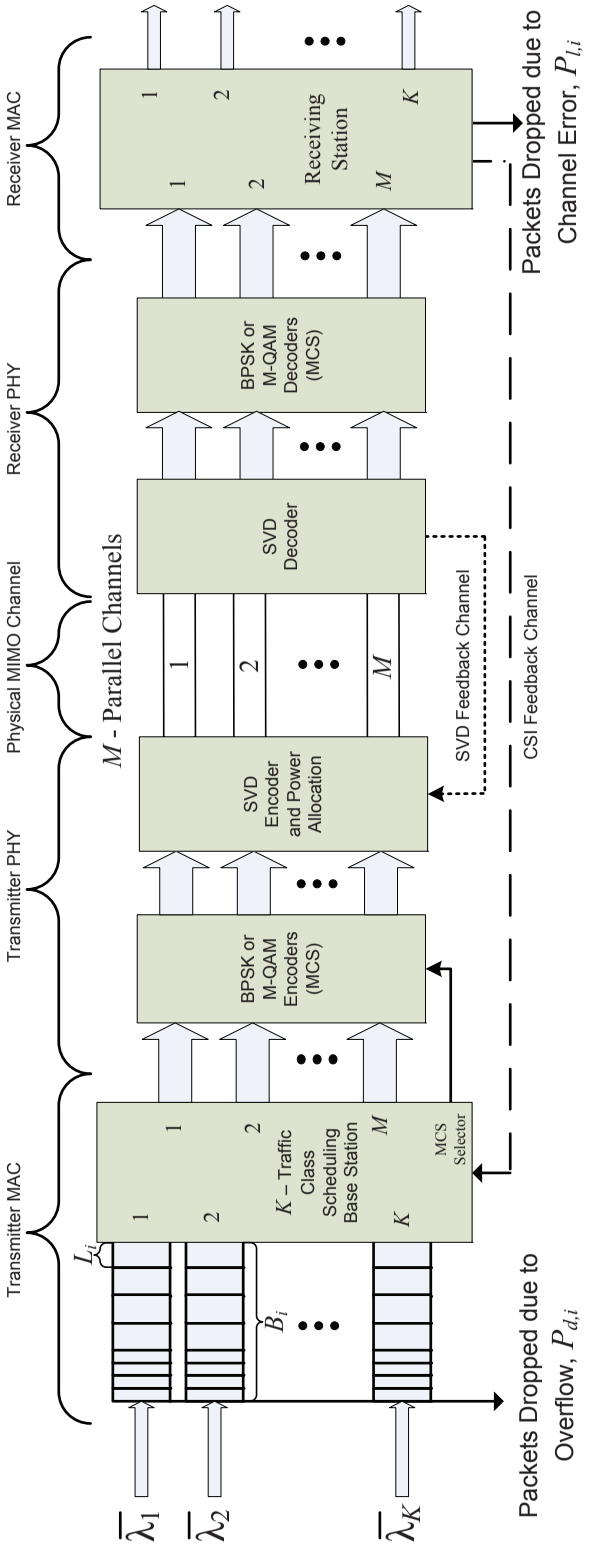


Figure 3.1: Downlink Multi-Queue, SVD MIMO Cross-Layer Model

3.1.1 MAC Layer Model

The media access control (MAC) layer model used in this work is as follows. Consider the downlink system shown in Figure 3.1. Traffic is received from upper layers and classified into K traffic streams. A single traffic stream has an associated set of QoS parameters $\{D_i, L_i, \overline{\lambda}_i, B_i, \delta_i\}$ which denotes the maximum tolerable average delay, fixed packet length, average arrival rate, buffer size and maximum tolerable packet loss rate respectively for that stream. Each stream may represent a broad service class (such as voice over IP or video) or a particular application-layer stream. Each incoming stream is stored in a finite-length first-in, first-out (FIFO) buffer where incoming packets are dropped when the buffer is full.

The probability of packet loss in the system (δ_i) is comprised of packets dropped at the source (due to buffer overflow, $P_{d,i}$) and packets dropped at the destination (due to channel errors, $P_{l,i}$). The total probability of an erroneous packet entering queue i is then $\delta_i = 1 - (1 - P_{d,i})(1 - P_{l,i})$. It is assumed that target probability of both loss types is known a priori.

The time horizon is divided into fixed scheduling intervals denoted as frames. Each frame has a duration of T_f seconds. Frame n is denoted as the interval of time bounded by $[nT_f, (n+1)T_f)$.

The K buffers are statistically multiplexed into a QoS-aware K-queue scheduler which makes scheduling decisions based on the CSI feedback from the subscribing station, the number of packets in each of the K MAC buffers, and their parameter set. As discussed in the previous chapter, while the theoretical number of queues (*i.e.,* K) with varying QoS constraints is large, practical implementations employ a finite number of classes [11] as most multimedia services can be categorized into one of several QoS classes.

3.1.2 PHY Layer Model

The PHY layer transmits packets scheduled for transmission during each frame. Packets are separated into M parallel streams and encoded with BPSK or M-QAM, where the constellation is decided based on channel conditions and MAC rate demands.

The M streams are then reconstructed at the receiver and forwarded to the receiver MAC. The M streams are provided by MIMO singular value decomposition (SVD)[1] achieved by assuming full channel knowledge is available at the transmitter error-free. The state of the M channels is characterized by their ordered eigenvalues where $\lambda_1^2 \geq \lambda_2^2 \geq \cdots \geq \lambda_M^2 \geq 0$. In general, MIMO SVD can provide $M \leq \min(M_T, M_R)$ streams (equal to the number of non-zero eigenvalues), however we assume that M is known and all eigenvalues are greater than 0. For simplicity at this stage, we also assume these eigenvalues are known, and do not change for all time (during the time interval $(0, \infty)$.

In general, the maximum number of parallel channels is equal to the minimum number of antennas at either the transmitting or receiving station. Recent measurement campaigns conducted in urban environments however suggest [55,56] that there are only a finite number of resolvable non-zero eigenvalues, which in general, can be less than the number of antennas.

Each independent channel is subject to noise. In the presence of additive white Gaussian noise (AWGN), the bit error rate of an uncoded M-ary signal is approximately [57]

$$Pb(\gamma_j, M_j) \approx 0.15 \exp\left(\frac{-1.55\gamma_j}{M_j - 1}\right), j = 1, 2, \ldots, M \tag{3.1}$$

where M_j is the size of the constellation set used in channel j, γ_j is the per symbol SNR given as $\gamma_j = P_j \gamma_0 \lambda_j^2$, P_j is the power allocated to channel j and γ_0 is the reference SNR level.

As discussed later, there is no constraint on the maximum instantaneous power selection, however we focus on minimizing the long-term average power consumption. The implications of this are discussed are discussed in Section 3.5.3.

1. The framework can easily be extended to any multi-channel system with known sub-channel error performance.

3.1.3 System Operation

From frame $n-1$ to frame n, the evolution of each buffer i follows

$$u_i(n) = \min\{B_i, \max\{0, u_i(n-1) - c_i(n)\} + A_i(n)\} \tag{3.2}$$

where $u_i(n)$ describes the buffer occupancy (number of buffer spaces in use) at the end of time frame n, B_i denotes the maximum buffer occupancy, $c_i(n)$ denotes the number of transmitted packets during the frame n (*i.e.,* the transmission action taken by queue i) and $A_i(n)$ denotes the number of arrivals in the queue. The queue evolution is shown pictorially in Figure 3.2. Figure 3.2a is the state of the queue at the beginning of frame n, Figure 3.2b is the arrivals to the queue during frame n and Figure 3.2c is the state of the queue at the end of frame n. Packets arriving during frame n (*i.e.,* $A_i(n)$) are assumed to be enqueued at the end of the frame. The quantities $c_i(n), u_i(n), A_i(n) \geq 0, \forall n$ and $c_i(n), u_i(n), A_i(n) \in \mathbb{I}$ where $\mathbb{I}$ is the set of all integers. The number of arrivals during frame n to a given queue (or $A_i(n)$) is a Poisson process with an average arrival rate of $\bar{\lambda_i}$ and a constant packet length of L_i. Further to this, packet arrivals are assumed to be independent of the current queue occupancy, service process and arrivals to other queues. For a Poisson process, the probability of k packets arriving to queue i during a frame of duration T_f is well-known to be

$$Pr[A_i(T_f) = k] = \begin{cases} \frac{(\bar{\lambda_i} T_f)^k \exp(-\bar{\lambda_i} T_f)}{k!}, & \text{if } k \geq 0 \\ 0, & \text{otherwise} \end{cases} \tag{3.3}$$

Packets are serviced in FIFO discipline over the previously described MIMO physical layer. In each individual parallel channel, bits are encoded from a finite modulation and coding scheme (MCS) alphabet $\mathcal{M}$ which determines the number of bits that can be encoded onto a single symbol. The selection of the set is described in later sections. Denoting k_j as the spectral efficiency[2] in bits/symbol for choosing

2. Spectral efficiency in this chapter refers to the number bits/symbol that can be transmitted in a given symbol duration, *i.e.,* 1 for BPSK, 2 for QPSK, and so on.

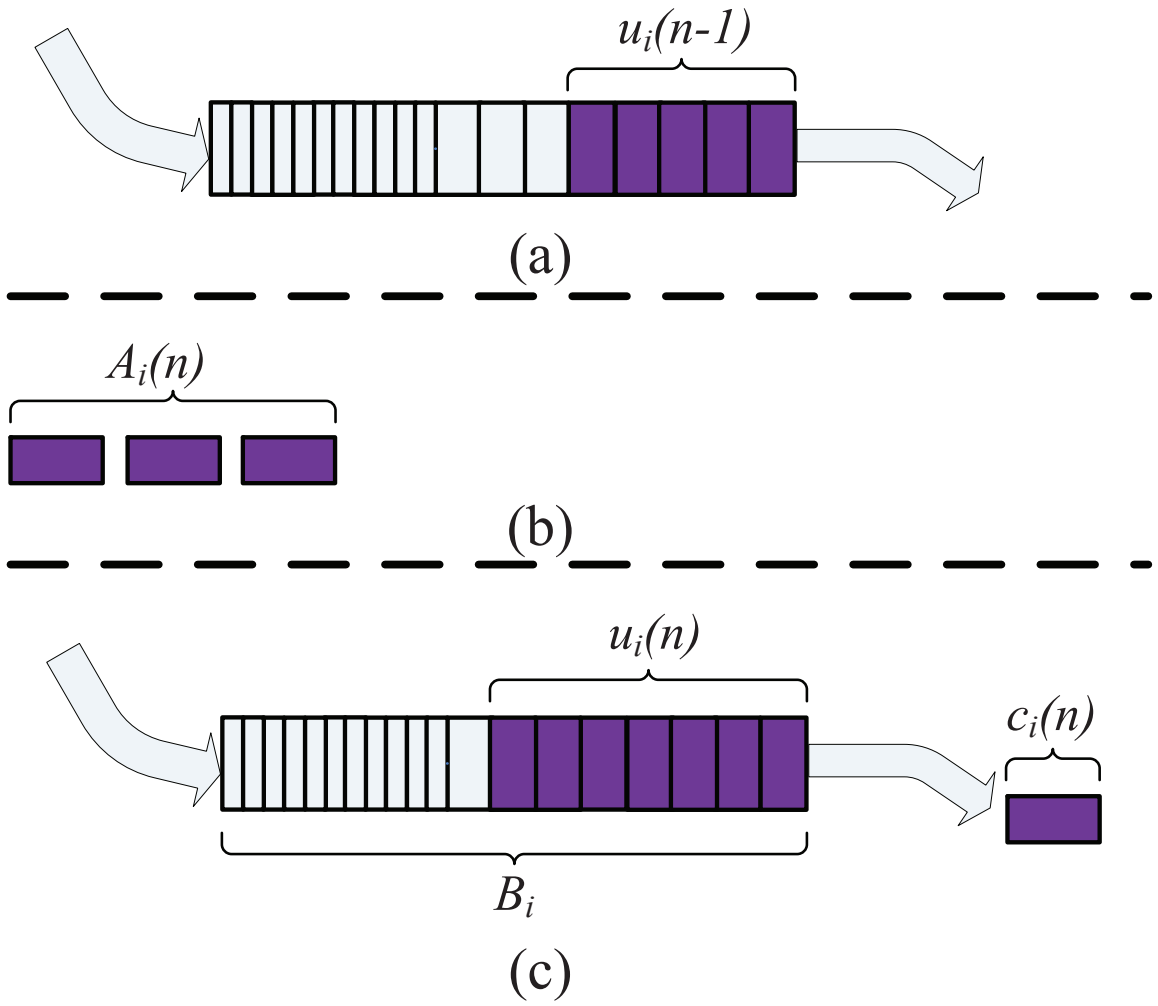


Figure 3.2: Timing Diagram of Queue Evolution

a constellation of size $M_j \in \mathcal{M}$ for the j^{th} channel and T_s as the symbol duration, the maximum number of bits that can be transmitted through the j^{th} channel over a duration of T_f seconds is $\frac{k_j T_f}{T_s}$ with a bit error rate given in (3.1).

3.2 $\{K \times M\}$ Scheduler Design

The proposed scheduler utilizes queue state information (QSI) to design a scheduling policy $\mathbf{\Omega}$. A given scheduling policy $\mathbf{\Omega}$ describes the channel, power and rate assignment for all time frames n. The benefit of utilizing QSI has been well demonstrated

by [20] and [21] in the case of a $\{1\times1\}$ system, however extensions for general $\{K\times M\}$ systems are non-trivial due to the complexity scaling and joint consideration of queue and channel state information. However, consider the following practical observation.

Examining only the buffer states for a set of K queues, the global set of states to describe the joint occupancy level across all queues is a K-dimensional set spanning the possible occupancy levels for the model discussed above and can be expressed as $\mathcal{U} := \mathcal{U}_1 \times \cdots \times \mathcal{U}_K$ where $\mathcal{U}_i = \{0, 1, \ldots, B_i\}$ is the state set for any buffer i and where 0 denotes an empty buffer

It was previously noted that arrivals to each queue are independent, however the service process couples the queues. In practice however, the number of possible service rates is comprised of a finite subset of the possible buffer states, and in our work physically represents the number of packets taken from the queue during a given frame. We previously denoted this quantity as $c_i(n)$ (*i.e.,* the transmission action taken by queue i during frame n). Now suppose, we have a set of rates which we denote $\mathcal{C}_i$ as the set of possible MAC rates (measured in packets) that can be serviced during each frame for queue i (*i.e.,* all possible values that can be taken on by $c_i(n)$, $\forall n$). This set is independent of the current frame n and we assume that $\mathcal{C}_i$ is chosen such that the maximum simultaneous packet transmission rate is achievable (*i.e.,* admission control is performed in advance to ensure the maximum transmission rate of all streams is less than the maximum rate on the channel). The overall MAC rate state-space of transmission actions is by extension simply $\mathcal{C} := \mathcal{C}_1 \times \cdots \times \mathcal{C}_K$. Assuming that $|\mathcal{C}_i| << |\mathcal{U}_i|$ which we argue occurs in practice due to various physical limitations such as employed symbol rates and scheduling intervals. By trivial extension is it is easily shown that $|\mathcal{C}| << |\mathcal{U}|$, where $|\cdot|$ denotes the size of a set. The preceding implies that any channel mapping and power control scheme need only consider $|\mathcal{C}|$ possible MAC layer rates rather than $|\mathcal{U}|$ possible states as a method of reducing the system complexity.

Using the above arguments, given a predetermined number of MAC service rates $\mathcal{C}_i$ for each queue i, we propose that the design of a $\{K \times M\}$ scheduler can be constructed as two components:

A: A mechanism to determine how to map a set of packets during each frame

across all parallel channels while performing rate and power adaptation to minimize power usage and maintain channel error rate requirements. This is computed for each $c \in \mathcal{C}$, where $\mathcal{C}$ is the MAC layer rate state-space, and

B: A mechanism to select the appropriate MAC layer transmission rate from each queue during each time frame to minimize total average transmission power found for $c \in \mathcal{C}$ in addition to ensuring QoS constraints are met by exploiting QSI.

The above problem segmentation allows the power, rate and channel allocation problems to be solved with reduced complexity by considering only a subset of information in each stage. As a result of the segmentation, components comprising the total average packet losses (*i.e.,* dropping probability and channel error probability) are no longer jointly optimized, and must be constrained individually at each stage.

3.2.1 Channel Mapping and Power Control

The first component of the scheduling mechanism maps a set of packets during a frame over the set of parallel channels. This procedure is performed for each possible MAC layer rate combination (*i.e.,* each $c \in \mathcal{C}$). The resulting outputs of this component are:

- A bit loading map $X_{j,i}(c), \forall j, i$ which denotes the number of bits per queue i that are mapped to channel j during the time frame,
- Constellation selection for each channel $(k_j(c), \forall j)$, and
- Power assignment for each channel $(P_j(c), \forall j)$

where $c \in \mathcal{C}$. Relevant complexity issues will be addressed in a later section. The constraints applied in this component are on packet losses on the channel, symbol rate selection and channel allocation. We note here that we do not constrain instantaneous power directly but allocated power is a function of the objective function. The following sections details the constraints applied in this component.

3.2.1.1 Packet Loss Due to Channel Error Constraint

Firstly, the average packet error rate on the channel experienced by a packet from stream i in state c given as

$$PER_i(c) = 1 - \prod_{j=1}^{M} (1 - P_{b,j}(c))^{x_{j,i}(c)L_i}, \forall i \tag{3.4}$$

where $P_{b,j}(c)$ is the bit error rate for channel j with power $P_j(c)$ assigned, spectral efficiency of $k_j(c)$ during state c given from (3.1) as

$$P_{b,j}(c) = Pb(P_j(c)\gamma_0\lambda_j^2, 2^{k_j(c)}) \tag{3.5}$$

and $x_{j,i}(c)$ is the percentage of bits per packet transmitted in the channel j from queue i such that $\sum_{j=1}^{M} x_{j,i}(c) = 1$.

Next, we assume the average packet error rate is targeted at each instant of time (*i.e.,* $PER_i(c) = \mathbb{E}_c[PER_i(c)]$ *where* $\mathbb{E}_c[\cdot]$ *is the expectation over* $\mathcal{C}$*).* The constraint on channel losses is then

$$1 - \prod_{j=1}^{M} (1 - P_{b,j}(c))^{x_{j,i}(c)L_i} \leq P_{l,i}, \quad \forall c, i \tag{3.6}$$

where $P_{l,i}$ is the target channel loss rate and is a portion of the total loss rate.

To further simplify (3.6), it can be approximated as shown in Appendix C as

$$1 - \left(1 - \sum_{j=1}^{M} x_{j,i}(c) P_{b,j}(c)\right)^{L_i} \leq P_{l,i}, \forall i \tag{3.7}$$

The above approximation assumes that the instantaneous bit error rate is $P_{b,j}(c) << 1$ which is true for target constraints on channel loss. Next, defining $X_{j,i}(c)$ as the number of bits mapped to channel j from queue i over a frame of duration T_f. Relating this to the quantity $x_{j,i}(c)$ discussed above we have

$$X_{j,i}(c) = x_{j,i}(c)c_i(c)L_i = x_{j,i}(c)R_i(c) \tag{3.8}$$

where $c_i(c)$ is defined as before in number of packets taken from queue i in state c and $R_i(c)$ is this quantity measured in bits. From (3.7), the final expression for the packet loss constraint is

$$P_{l,i} \geq 1 - \left(1 - \sum_{j=1}^{M} \frac{X_{j,i}(c) P_{b,j}(c)}{R_i(c)}\right)^{L_i}, \forall i \tag{3.9}$$

Which can be constrained in terms of the bit error rates such that

$$\sum_{j=1}^{M} \frac{X_{j,i}(c) P_{b,j}(c)}{R_i(c)} \leq BER_i, \forall i, \tag{3.10}$$

where $BER_i = 1 - (1 - P_{l,i})^{\frac{1}{L_i}}$

3.2.1.2 Rate Selection Constraints

We also select constellation schemes for each channel such that the requested MAC layer rate requirement is met. The total MAC layer rate requirement is given as $\sum_{i=1}^{K} R_i(c)$. Therefore we note that

$$\sum_{j=1}^{M} \frac{k_j(c) T_f}{T_s} \geq \sum_{i=1}^{K} R_i(c) \tag{3.11}$$

is a necessary condition. We also note that the set $\{k_j\}_{j=1}^{M}$ that satisfies the above such that $\sum_{j=1}^{M} k_j$ should be minimized to achieve the minimum power usage.

Since by design $k_j \in \mathcal{M}$ is a discrete set and (3.1) is monotonic in k_j, one can easily show that any set $\{k_j\} \in \mathcal{M}$ that minimizes transmission power must satisfy

$$\sum_{j=1}^{M} k_j(c) = \left\lceil \frac{T_s}{T_f} \sum_{i=1}^{K} R_i(c) \right\rceil \tag{3.12}$$

where in this case, $\lceil \cdot \rceil$ denotes rounding up to the nearest valid sum of k_j values. The set of all valid MCS mode combinations that satisfies the above for each $c \in \mathcal{C}$

is denoted $\mathcal{K}_{min}(c)$.

While (3.12) is true if valid spectral efficiencies in each channel increment by 1, the values that can be taken on by k_j fall within the set of allowable MCS modes and in this work is further restricted to $\{0, 1, 2, 4, 6\}$ representing no transmission, BPSK, QPSK and 16/64-QAM respectively. Therefore a marginal increase in transmission rate may result in an minimum increase of spectral efficiency of 2 in one channel. To integrate this phenomenon, the search set of valid MCS modes is extended such that

$$\mathcal{K}(c) = \begin{cases} \mathcal{K}_{min}(c) \bigcup \mathcal{K}_{min+1}(c), & \sum_{i=1}^{K} R_i > 0 \\ 0 & otherwise \end{cases} \tag{3.13}$$

where $\mathcal{K}_{min+1}(c)$ is the set of MCS modes satisfying $\left\lceil \frac{T_s}{T_f} \sum_{i=1}^{K} R_i(c) \right\rceil + 1$.

3.2.1.3 Channel Mapping Constraints

Based on the system design, we also have the following constraints on the mapping coefficients $X_{j,i}(c)$:

$$\sum_{j=1}^{M} X_{j,i}(c) = R_i(c), i = 1, 2, \ldots K \tag{3.14}$$

$$\sum_{i=1}^{K} X_{j,i}(c) \leq \frac{k_j(c) T_f}{T_s}, j = 1, 2, \ldots, M \tag{3.15}$$

3.2.1.4 Power Control

Minimization of the average applied power is the objective function. Average power is the sum of power allocated in each subchannel multiplied by the number of symbols transmitted over that subchannel or simply

$$\sum_{j=1}^{M} \frac{P_j(c)}{k_j(c)} \sum_{i=1}^{K} X_{j,i}(c) \tag{3.16}$$

The above set of constraints, which contains both integer and discrete variables can be formulated as a mixed-integer non-linear programming (MINLP) problem. The details of the MINLP formulation and elements are discussed in Section 3.3.1.

3.2.2 Locally Optimized MAC Rate Selection

The second component of the $\{K \times M\}$ scheduler design is to select a MAC layer transmission rate (or $c_i \in \mathcal{C}_i$) to determine the number of packets to transmit during frame n from queue i. This decision is based on both the QoS constraints and the power allocation determined by (3.16).

Each queue is characterized by its current state $u_i \in \mathcal{U}_i$ denoting the current occupancy level. During anytime n, c_i packets may be taken from the queue where $\mathcal{C}_i$ is the set of all transmission actions (packets that can be transmitted) during a given frame. The scheduling policy $\mathbf{\Omega}$ defines the set probabilities of choosing c_i when the current queue state is u_i for each queue i. From (3.2) it can be seen that the queue occupancy during any frame n depends only on the occupancy during frame $n-1$ and arrivals during that frame. As such, the above can be solved as constrained Markov decision process (CMDP) [51] to obtain a scheduling policy $\mathbf{\Omega}$. Let $\theta_i(c_i, u_i|\mathbf{\Omega})$ be a steady-state distribution function that exists for a particular policy $\mathbf{\Omega}$ which denotes the probability of being in state u_i and transmitting c_i packets during frame n. The scheduling policy $\mathbf{\Omega}$ is obtained through application of the constraints on average delay and MAC layer throughput given as follows.

3.2.2.1 Throughput Constraint

The dropping probability is related to the MAC throughput by

$$\overline{\chi}_i = \lambda_i(1 - P_{d,i})T_f \tag{3.17}$$

Therefore we do not constrain the dropping probability directly, but rather constrain the minimum MAC layer throughput. The throughput at each state u_i is dependant on both the queue state and the action taken (*i.e.,* u_i and c_i). For a given set $\{c_i, u_i\}$

during frame n, the throughput is

$$\chi_{i:n}(c_i, u_i) = \min(c_i, u_i) \tag{3.18}$$

3.2.2.2 Delay Constraint

From Little's Theorem, the average queueing delay constraint is

$$D_i \geq \mathcal{D}_i = \frac{\bar{q}_i}{\lambda_{q,i} T_f} \tag{3.19}$$

where $\bar{q}_i$ is the average queue size and $\lambda_{q,i}$ is the average enqueued arrival rate for queue i. By design we can express $\bar{q}_i$ using the steady-state distribution $\theta_i(c_i, u_i|\mathbf{\Omega})$ as:

$$\bar{q}_i = \sum_{u_i \in \mathcal{U}_i} u_i \sum_{c_i \in \mathcal{C}_i} \theta_i(c_i, u_i|\mathbf{\Omega}) \tag{3.20}$$

and since $\lambda_{q,i}$ is also equal to the average service rate in steady-state, it can be expressed as

$$\lambda_{q,i} = \sum_{u_i \in \mathcal{U}_i} \sum_{c_i \in \mathcal{C}_i} \min(c_i, u_i) \theta_i(c_i, u_i|\mathbf{\Omega}) \tag{3.21}$$

3.2.2.3 Transition Probabilities

The transition probabilities denote the probability of transitioning from one queue state to another. By design this is based on the arrival process, the given state, the next state and the transmission action taken. With all quantities defined as before, we denote $p^{c_i}_{u_i;u'_i}$ as the probability of transitioning from state u_i to u'_i given action c_i is taken. From (3.2) and (3.3) this is given as

$$p^{c_i}_{u_i;u'_i} = \begin{cases} P(A_i(T_f) = u'_i - [u_i - \min(u_i, c_i)]), & u'_i < B_i \\ \sum_{j=B_i-[u_i-\min(u_i,c_i)]}^{\infty} P(A_i(T_f) = j), & u'_i = B_i \end{cases} \tag{3.22}$$

By design, the steady-state distribution $\theta_i(c_i, u_i|\mathbf{\Omega})$ must also satisfy the fol-

lowing balance property

$$\sum_{u_i' \in \mathcal{U}_i} \sum_{c_i' \in \mathcal{C}_i} \theta(c_i', u_i'|\mathbf{\Omega}) p^{c_i'}_{u_i';u_i} = \sum_{c_i \in \mathcal{C}_i} \theta(c_i, u_i|\mathbf{\Omega}), \forall u_i \tag{3.23}$$

3.2.3 Per Queue Objective Function

The power allocation found in the first component is used to derive the objective function for the local MAC layer rate assignment. First, the average marginal cost for taking an action c_i in queue 1 can be given as

$$\Upsilon_{1,x} = \sum_{c_2 \in \mathcal{C}_2} \dots \sum_{c_K \in \mathcal{C}_K} P(x, c_2, \dots, c_K) \cdot \pi_2(c_2|\mathbf{\Omega}) \times \dots \times \pi_K(c_K|\mathbf{\Omega}) \tag{3.24}$$

where there are $i-1$ summations. Similar expressions can be found for all actions $c_i \in \mathcal{C}_i$ and found for all queues $k = 1, \dots, K$ and where

$$\pi_i(x|\mathbf{\Omega}) = \sum_{u_i \in \mathcal{U}_i} \theta(x, u_i|\mathbf{\Omega}), x \in \mathcal{C}_i \tag{3.25}$$

The above steady-state action probabilities are coupled through the policy $\mathbf{\Omega}$. The value $P(c_1, c_2, \dots, c_K)$ is the total power associated with taking actions c_1 through c_K in each queue (or one for each state $c \in \mathcal{C}$) found as the solution to (3.16). Here we need to highlight that the above expression contains the steady-state probability of choosing an action in each queue. The result of which implies that it is not possible to directly decouple and consider each queue independently. We can however consider the following special cases.

1) *Single Queue:* For the single queue case, the cost function in (3.24) reduces to the total power required and can be solved as an linear programming (LP) problem.
2) *Two Queues:* For the two queue scenario, the cost function model can be considered a Quadratic Programming (QP) problem where the number of degrees of freedom is twice that of the single queue problem or $|\mathcal{C}_1 \times \mathcal{U}_1| + |\mathcal{C}_2 \times \mathcal{U}_2|$, rather than $|\mathcal{C}_1 \times \mathcal{U}_1 \times \mathcal{C}_2 \times \mathcal{U}_2|$.

3) General Number of Queues: In general, the problem can be solved using iterative methods. This process is as follows. Firstly, given the power allocation values, we iteratively solve the C-MDP problem (as an LP problem) for each queue and update the corresponding cost function until the steady state distribution $\theta_i(c_i, u_i|\mathbf{\Omega})$ in each queue converges. Convergence details are discussed in Section 3.4.3.

A summary of frequently used notation is provided in Table 3.1.

3.3 Formulating the Problem Using Optimization Framework

Both the channel/power allocation and the local MAC rate assignment mechanisms are formulated as optimization problems using framework described in Appendix B. The channel and power allocation scheme can be formulated as a generic MINLP problem, while the local MAC rate assignment can be formulated as a general LP (or QP) problem.

3.3.1 Formation of MINLP Problem

A general solution to a NLP problem is non-trivial, this is further complicated by introduction of discrete or integer constraints on several variables. To relax these discrete constraints we perform the following:

1) Consider $X_{j,i}(c)$ as a continuous variable as we note in practice rounding $X_{j,i}$ to the nearest integer affects only a single bit of information. Since during any frame where there is an active the transmission the number of transmitted bits is in general much larger than 1, rounding does not dramatically affect the result, and

2) We formulate a general NLP problem for each subset satisfying (3.12) and choose the allocation strategy achieving the lowest power consumption.

Based on the above, we formulate a general NLP problem such that we solve $\arg\min_{\mathbf{x}} f(\mathbf{x})$ subject to $\mathbf{A}\mathbf{x} \leq \mathbf{b}$, $\mathbf{A}_{eq}\mathbf{x}_{eq} = \mathbf{b}_{eq}$, $\mathbf{x} \geq 0$ and $\mathbf{c}(\mathbf{x}) \leq 0$ where $\mathbf{A}$

Table 3.1: Frequently Used Notation Used Throughout This Chapter

Quantity	Symbol	Quantity	Symbol
Number of Traffic Streams	K	Subchannel SNR	γ_j
Number of Parallel Channels	M	BER of Channel j	$P_{b,j}$
Average Delay Constraint	$\mathcal{D}_i$	Scheduling Policy	$\mathbf{\Omega}$
Packet Size in Bytes	L_i	Buffer state-space	$\mathcal{U}_i$
Average Arrival Rate	$\bar{\lambda}_i$	MAC Rate state-space	$\mathcal{C}_i$
Buffer Size	B_i	Joint MAC Rate state-space	$\mathcal{C}$
Total Average Loss Constraint	δ_i	Transition probability	$p^{c_i}_{u_i,u_i'}$
Packet Dropping Probability	$P_{d,i}$	Fraction of bits allocated to channel j from stream i	$x_{j,i}$
Probability of Channel Packet Loss	$P_{l,i}$	Number of bits allocated to channel j from stream i	$X_{j,i}$
Frame Duration	T_f	Stream Rate of channel i	R_i
Symbol Duration	T_s	Throughput of stream i	$\bar{\chi}_i$
Frame Number	n	Per queue cost function	$\Upsilon_{i,x}$
Subchannel Eigenvalue	λ_j^2	Steady-State policy distribution	$\theta_i(c_i, u_i\|\mathbf{\Omega})$
Reference SNR	γ_0	Steady-state action probability	$\pi_i(x\|\mathbf{\Omega})$
Set of Valid MCS Modes	$\mathcal{M}$	Number of Arrivals during frame n	$A_i(n)$
Spectral efficiency in channel j	k_j	Buffer Occupancy during frame n	$u_i(n)$
M-ary Mode	M_j	Packet Service rate of queue i during frame n	$c_i(n)$

and $\mathbf{A}_{eq}$ are matrices, $\mathbf{b}$ and $\mathbf{b}_{eq}$ are vectors, $\mathbf{c}(\mathbf{x})$ is a vector of non-linear functions evaluated at $\mathbf{x}$ and $f(\mathbf{x})$ is a scalar non-linear function evaluated at $\mathbf{x}$. The above is evaluated at each state $c \in \mathcal{C}$ and over the space $\mathcal{K}(c)$ which is the space containing each combination $\{k_j\}_{j=1}^{M}$ that meets the rate selection restrictions above described in (3.12).

The derivations of the NLP elements is given below. The vector $\mathbf{x}$ is a $(M + MK) \times 1$ vector with elements $P_j, j = 1, 2, \ldots, M$ and $X_{j,i}, j = 1, 2, \ldots, M, i = 1, 2, \ldots, K$ given as

$$\mathbf{x} = [P_1, \ \ldots, \ P_M, \ X_{1,1}, \ \ldots, \ X_{1,K}, \ \ldots, \ X_{M,K}]^T \tag{3.26}$$

3.3.1.1 Objective Function

The objective function from $f(\mathbf{x})$ given in (3.16) is formulated as

$$f(\mathbf{x}) = \sum_{j=1}^{M} \frac{\mathbf{x}(\mathcal{P}_j)}{k_j} \sum_{i \in \mathcal{I}'_j} \mathbf{x}(i) \tag{3.27}$$

where $\mathcal{I}'_j$ and $\mathcal{P}_j$ are the sets containing location indices of $X_{j,i}, \forall i$ and P_j respectively in $\mathbf{x}$.

3.3.1.2 Equality Constraints

The K equality constraints from (3.14) are given in the $K \times (M + MK)$ matrix $\mathbf{A}_{eq}$ with entries

$$A_{eq:i,k} = \begin{cases} 1, k \in \mathcal{I}_i \\ 0, otherwise \end{cases} \tag{3.28}$$

where $\mathcal{I}_i$ is the set containing location indices of $X_{j,i}, \forall j$ in $\mathbf{x}$. The coefficient vector $\mathbf{b}_{eq}$ is given as

$$\mathbf{b}_{eq} = [R_1, \ R_2, \ldots, \ R_K]^T \tag{3.29}$$

3.3.1.3 Inequality Constraints

The M equality constraints from (3.15) are defined in the $M \times (M + MK)$ matrix $\mathbf{A}$ with entries

$$A_{j,k} = \begin{cases} 1, k \in \mathcal{I}'_j \\ 0, otherwise \end{cases} \tag{3.30}$$

The coefficient vector $\mathbf{b}$ is given as

$$\mathbf{b} = \frac{T_f}{T_s}[k_1,\ k_2, \ldots,\ k_M]^T \tag{3.31}$$

3.3.1.4 Non-Linear Inequality Constraint

The K non-linear inequality constraints are given as a $K \times 1$ vector of functions of $\mathbf{x}$. For simplicity of notation with further define $\mathcal{I}_{j,i}$ as the indices of $X_{j,i}, \forall i, j$. Using the bit error rate expression in (3.1), we then have

$$\mathbf{c}(\mathbf{x}) = [c_1(\mathbf{x}),\ \ldots,\ c_K(\mathbf{x})]^T \tag{3.32}$$

where

$$c_i(\mathbf{x}) = \frac{\sum_{j=1}^{M} 0.15 \exp\left(\frac{-1.55\mathbf{x}(\mathcal{P}_j)\gamma_0\lambda_j^2}{(2^{k_j}-1)}\right)\mathbf{x}(\mathcal{I}_{j,i})}{R_i} - BER_i$$

The solution can be found for the above framework using general NLP methods such as those provided by *fmincon* included in the MATLAB optimization toolbox.

3.3.2 Forming LP Problem

As in [51], the constrained MDP problem formulated for the local MAC rate selection can be solved using Linear Programming (LP) techniques for each queue. LP techniques efficiently solve convex optimization problems of the form $\arg\min_{\mathbf{x}} \mathbf{c}^T\mathbf{x}$, subject to $\mathbf{A}\mathbf{x} \leq \mathbf{b}$, $\mathbf{A}_{eq}\mathbf{x} = \mathbf{b}_{eq}$, $\mathbf{x} \geq 0$ where $\mathbf{A}$ and $\mathbf{A}_{eq}$ are matrices and $\mathbf{x}, \mathbf{b}, \mathbf{b}_{eq}$ and $\mathbf{c}$ are column vectors. The vector $\mathbf{x}$ is the solution to the optimization problem. In our

problem, the elements are given as

$$\mathbf{x} = [\boldsymbol{\theta}_i(\mathcal{C}_i, 0|\boldsymbol{\Omega}), \ldots, \boldsymbol{\theta}_i(\mathcal{C}_i, B_i|\boldsymbol{\Omega})]^T \tag{3.33}$$

with each $\boldsymbol{\theta}_i(\mathcal{C}_i, u_i|\boldsymbol{\Omega})$ being a row vector with entries for each $c_i \in \mathcal{C}_i$.

3.3.2.1 Objective Function

The objective function is of the form $\mathbf{c}^T\mathbf{x}$. The vector $\mathbf{c}$ is comprised of the total power cost for taking an action. Each entry of $\mathbf{c}$ corresponds to the entry in $\mathbf{x}$ with the value of entries in $\mathbf{c}$ given by Υ_{i,c_i} in (3.24).

$$\mathbf{c} = [\underbrace{\Upsilon_{i,1}, \ldots, \Upsilon_{i,|\mathcal{C}_i|}}_{1}, \overbrace{\ldots, \ldots, \ldots}^{2..B_i}, \underbrace{\Upsilon_{i,1}, \ldots, \Upsilon_{i,|\mathcal{C}_i|}}_{B_i+1}] \tag{3.34}$$

3.3.2.2 Equality Constraints

The equality constraints are comprised of the balance equations and the causality constraint (total probability space) given in (3.22) and (3.23) respectively. In matrix form, the balance equations can be expressed as $\mathbf{P} \times \mathbf{x} = \boldsymbol{\Phi}_0 \times \mathbf{x}$ where $\mathbf{P}$ is given by

$$\mathbf{P} = \begin{bmatrix} \mathbf{p}_{0;0}^{\mathcal{C}_i} & \cdots & \cdots & \mathbf{p}_{B_i;0}^{\mathcal{C}_i} \\ \vdots & \mathbf{p}_{1;1}^{\mathcal{C}_i} & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{p}_{0;B_i}^{\mathcal{C}_i} & \cdots & \cdots & \mathbf{p}_{B_i;B_i}^{\mathcal{C}_i} \end{bmatrix} \tag{3.35}$$

with $\mathbf{p}_{q;q'}^{\mathcal{C}_i}$ as a $1 \times |\mathcal{C}_i|$ row vector with entries

$$\mathbf{p}_{q;q'}^{\mathcal{C}_i} = [p_{q;q'}^1, \ldots, p_{q;q'}^{|\mathcal{C}_i|}] \tag{3.36}$$

and the quantity $\mathbf{\Phi}_0$ is given as the $B_i + 1$ row matrix

$$\mathbf{\Phi}_0 = \begin{bmatrix} \mathbf{1}_{1\times|\mathcal{C}_i|} & 0 & \cdots & 0 \\ 0 & \mathbf{1}_{1\times|\mathcal{C}_i|} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{1}_{1\times|\mathcal{C}_i|} \end{bmatrix} \tag{3.37}$$

Combining the above with the causality constraint on the total probability space we have our overall equality constraints given as

$$\mathbf{A}_{eq} = \begin{bmatrix} \mathbf{P} - \mathbf{\Phi}_0 \\ \mathbf{1}_{1\times(|\mathcal{C}_i|(B_i+1))} \end{bmatrix} \qquad \mathbf{b}_{eq} = [\mathbf{0}_{1\times(B_i+1)} \quad 1]^T \tag{3.38}$$

3.3.2.3 Inequality Constraints

The inequality constraints are used to describe the throughput and delay constraints and are given by (3.17) and (3.19) respectively. These constraints are given in two parts as

$$\mathbf{A} = \begin{bmatrix} \mathbf{w}_1 \\ \mathbf{w}_2 \end{bmatrix} \qquad \mathbf{b} = \begin{bmatrix} \mathbf{z}_1 \\ \mathbf{z}_2 \end{bmatrix} \tag{3.39}$$

where $\mathbf{w}_1$ is given as

$$\mathbf{w}_1 = -[\boldsymbol{\chi}_{i:n}(\mathcal{C}_i, 0), \ldots, \boldsymbol{\chi}_{i:n}(\mathcal{C}_i, B_i)] \tag{3.40}$$

where $\boldsymbol{\chi}_{i:n}(\mathcal{C}_i, u_i)$ is a row vector with entries $\chi_{i:n}(c_i, u_i)$ for all $c_i \in \mathcal{C}_i$ and $\mathbf{z}_1$ is given as

$$\mathbf{z}_1 = -\bar{\lambda}_i(1 - P_{d,i})T_f \tag{3.41}$$

Further, using (3.19)-(3.21), and after some trivial manipulation we can obtain

$$\mathbf{w}_2 = \mathbf{Q} \times \mathbf{\Phi}_0 - \mathcal{D}_i\mathbf{U} \qquad \mathbf{z}_2 = 0 \tag{3.42}$$

where $\mathbf{Q} = [0, 1, \ldots, B_i]$ and $\mathbf{U}$ is given as

$$\mathbf{U} = [\min(c_i(1), 0), \ldots, \min(c_i(|\mathcal{C}_i|), 0), \min(c_i(1), 1), \ldots, \min(c_i(|\mathcal{C}_i|), B_i)] = -\mathbf{w}_1 \tag{3.43}$$

where $c_i(x)$ denotes the x^{th} MAC rate in $\mathcal{C}_i$.

The above framework can be computed using general LP methods such as those provided by *linprog* included in the MATLAB optimization toolbox for both the single queue and iterative methods. Extensions are trivial for the special case of two queues using QP methods.

The LP problem above yields the steady-state distribution $\theta_i(c_i, u_i|\mathbf{\Omega}), c_i \in \mathcal{C}_i, u_i \in \mathcal{U}_i, i \in \{1, 2, \ldots, K\}$. The physical meaning of this solution is that in a given queue i while the buffer is in a given state u_i, the scheduler selects c_i packets from the queue for transmission with probability given by

$$\Pr[c_i|u_i, i, \mathbf{\Omega}] = \frac{\theta_i(c_i, u_i|\mathbf{\Omega})}{\sum\limits_{c_i' \in \mathcal{C}_i} \theta_i(c_i', u_i|\mathbf{\Omega})} \tag{3.44}$$

3.3.3 Scheduler Implementation

The above optimization problem is solved offline, in advance with proper channel measurements. The advantage of this method is that all quantities can be stored in a lookup table (LUT). The LUT stores information on the power, MCS mode and bit allocation for each state $c \in \mathcal{C}$. The offline scheduler works as follows. At the beginning of each frame n, each queue has u_i packets waiting for transmission. All queues then choose actions c_i with probability given in (3.44). Given a joint action $c = \{c_i, \forall i\}$, the scheduler selects the stored MCS and power modes for each channel and allocates bits from all queues as found in the stored bit allocation.

All quantities in the LUT can be accurately stored as 64 bit double. The space required to store each state is

$$Size_c = 64(KM + 2M) \quad \text{bits/state} \tag{3.45}$$

as we require storage of KM bit allocations, M power levels and M MCS mode selections. Further, we note the system has $|\mathcal{C}|$ states, therefore the total size of the LUT in bits is given as

$$Size_{LUT} = |\mathcal{C}|Size_c = 64|\mathcal{C}|(KM + 2M) \quad \text{bits} \tag{3.46}$$

where the above describes the relation of number of channels, the number of queues and the possible MAC rates to the storage size of the LUT.

3.4 Simulation Results and Discussion

We provide simulation results for the three MAC layer rate assignment approaches. Firstly, results are provided for the special case of a single queue using the LP approach, followed by application of the QP approach for the case of two queues. Finally we validate the iterative approach by comparing the accuracy of the two queue system with QP approach. Convergence details of the iterative method are also provided.

Table 3.2: Simulation Parameters

Parameter	Value
Number of Antennas (M)	4
Number of Queues (K)	1
Spectral Efficiencies	$\{0,\ 1,\ 2,\ 4,\ 6\}$
Length of Packet (bits)	200
Arrival Rate (Packets/frame)	1
Queue Size (Packets)	25
Average Packet Delay (Frames)	5
Total Loss Rate (δ, % of Packets)	1%
Target Channel Loss Rate	$\delta/2$
Frame Duration (T_f)	1
Symbol Duration (T_s)	0.01
MAC Rates (Packets per Frame)	$\{0, 1, 2, 3, 4, 5\}$
MIMO Channel Eigenvalues	[2, 1.5, 0.6, 0.4]
Reference SNR (γ_0)	20dB

Further, we also provide some metrics for the system complexity. Universal and single queue simulation parameters are shown in Table 3.2 while parameters for the two queue system are given in Table 3.3.

The performance of the proposed scheme is compared with a strict scheduler. The strict scheduler implementation employs the same physical layer as the proposed scheme, however the transmission rate (packets per frame) is constant. This transmission rate is chosen from the set of allowable transmission rates such that it is the minimum rate that ensures QoS constraints (dropping probability and delay) are guaranteed and is found through a M/D/1/N model [58].

3.4.1 Single Queue Performance

Figures 3.3-3.6 show the results for total average power for the single queue scenario for varying of major parameters. In all cases the proposed scheduler requires lower average power for transmission. For the case of increasing arrival rate or packet size we note an increase in the power selection which is expected while in the case of delay, increasing delay constraints result in a reduction of power. Once the delay constraint exceeds a certain threshold, power is no longer reduced as the scheduler is no longer dominated by the delay constraint. In the case of increasing loss rate we see a steady decline in power. This is due to the large amount of power required to maintain the stringent loss requirements (by maintaining a low channel error rate and lower number of packets dropped in the buffer). Finally, power selection is u-shaped when the buffer size is varied. This variation is negligibly small relative to the difference

Table 3.3: Simulation Parameters for 2 Queue Scenario

Parameter	Value
Length of Packet (bits)	[200 250]
Arrival Rate (Packets/frame)	[1 1]
Queue Size (Packets)	[25 25]
Average Packet Delay (Frames)	[4 5]
Total Loss Rate (δ, % of Packets)	[1 1]%
Target Channel Loss Rate	$\delta/2$

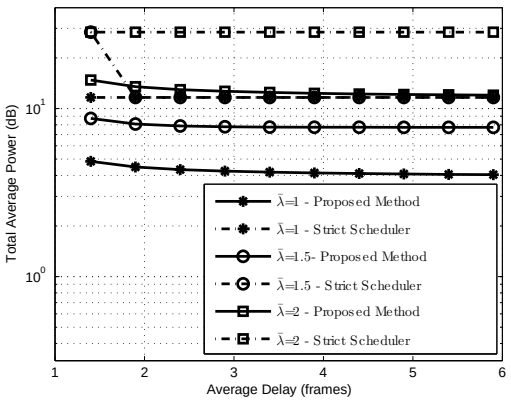


Figure 3.3: Total Average Power vs. Arrival Rate - Single Queue

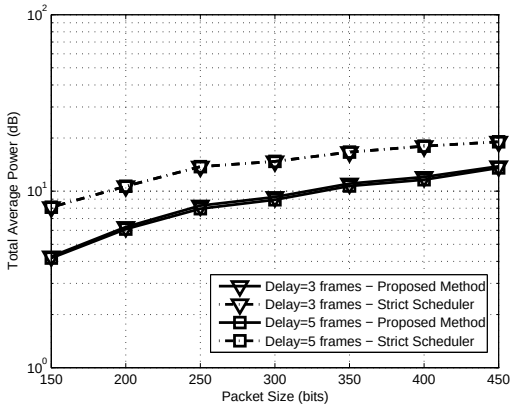


Figure 3.4: Total Average Power vs. Packet Size - Single Queue

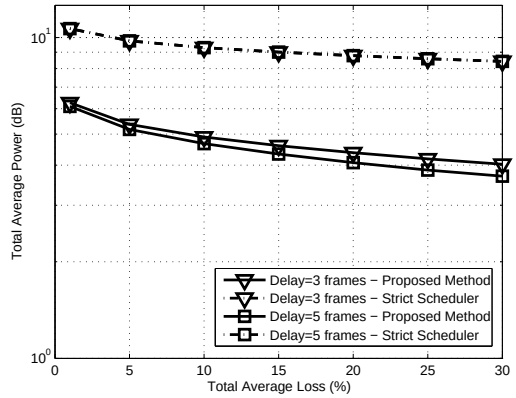


Figure 3.5: Total Average Power vs. Loss Rate - Single Queue

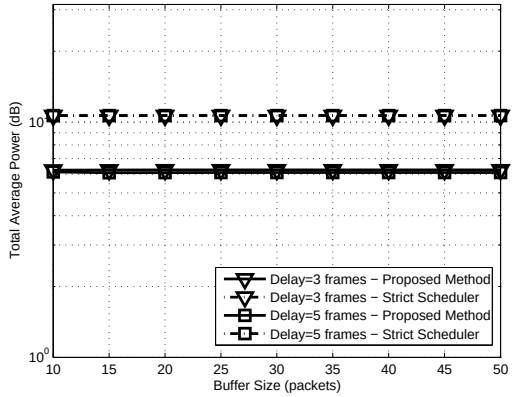


Figure 3.6: Total Average Power vs. Buffer Size - Single Queue

between the proposed method and the strict scheduler (shown in Figure 3.6). The explanation for this phenomenon is as follows. In the smaller buffer region, power selection is dominated by trying to maintain acceptable loss in the buffer while in the larger buffer region, the power selection hits a plateau due to the limitation in the maximum tolerable delay is no longer a dominant factor.

3.4.2 Two Queue Performance

In Figures 3.7-3.10 the results for the two queue scenario is shown. As expected, the trends for the two queue scenario follow those for the case of a single queue

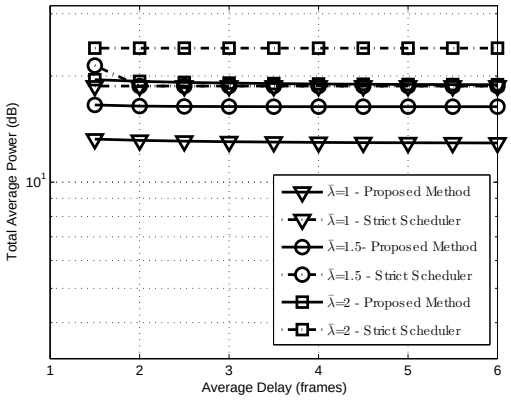


Figure 3.7: Total Average Power vs. Arrival Rate - Two Queues

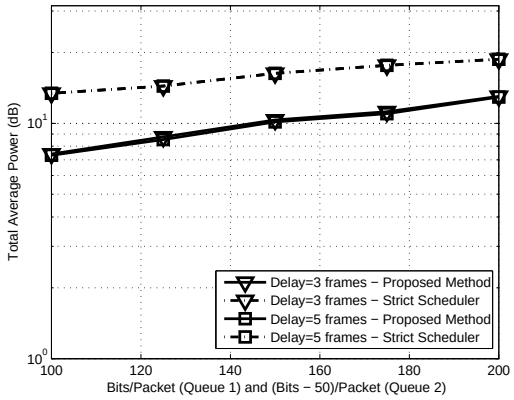


Figure 3.8: Total Average Power vs. Packet Size - Two Queues

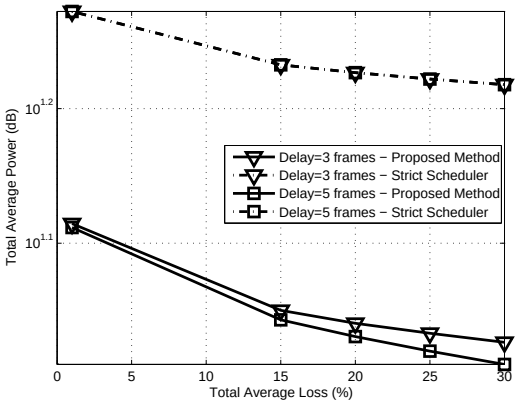


Figure 3.9: Total Average Power vs. Target Loss - Two Queues

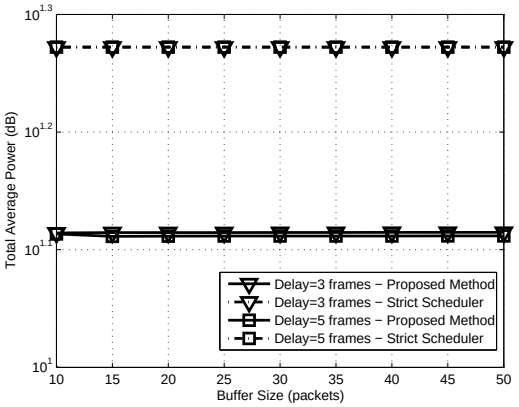


Figure 3.10: Total Average Power vs. Buffer Size - Two Queues

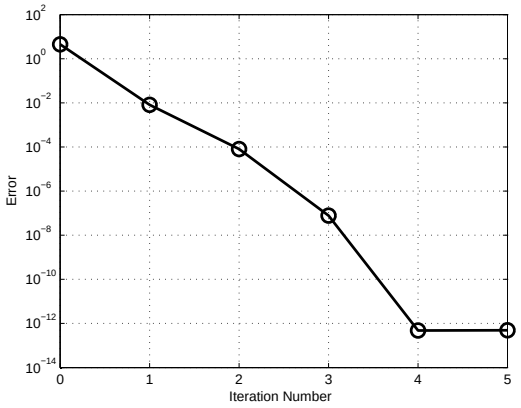


Figure 3.11: Relative Error of Iterative Policy Solver vs. Number of Iterations

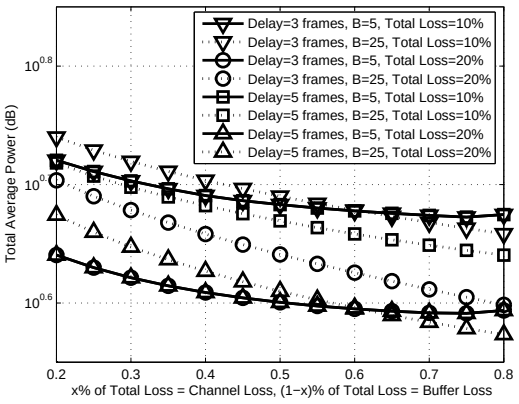


Figure 3.12: Impact of Varying Target Loss Rate Component

with higher transmission power. This is due to the increase in the total required physical layer transmission rate. Further, for the case of two queues, we utilize both the iterative and QP methods to validate its convergence. Our proposed iterative method converges to the same solution as the QP approach. Further, the storage requirements of the LUT in this case is only $36,864$ bits (or 4.5 kilobytes).

3.4.3 Iterative Convergence Discussion

For 3 or more queues, it is necessary to utilize the iterative LP method where each queue is solved as a LP problem and updates the cost function in (3.24). This method is guaranteed to find a local minimum due to the monotonicity of (3.24) in c_i irrespective of the values of $\pi_{i'}(c_{i'}|\mathbf{\Omega})$ for $\forall i', i' \neq i$ and due to the convexity of the LP problem within each queue.

We further demonstrate the global convergence using Monte Carlo simulations for both 2 and 3 queue scenarios over 10000 random initial solutions. For the two queue scenario, results are compared with the QP result, and with the 3 queue scenario results are measured as relative error (deviation from minimal solution). The average relative error versus iteration number in this case is shown in Figure 3.11. As shown, the iterative solution has a maximal deviation of less than 10^{-7} over 10000 random realizations after just 3 iterations strongly suggesting that the iterative method converges to the global minimum.

3.4.4 Loss Constraint Related Tradeoffs

As previously discussed, we briefly overview the tradeoff between buffer and channel losses on the energy efficiency of the proposed design since joint optimization of these losses is a drawback of our proposed framework. Figure 3.12 shows the total average power versus the percentage of the total loss rate for both the buffer and channel losses. This is given for a several configurations of total loss rates, average delay constraints and buffer sizes. Overall, the trends demonstrate that is more efficient to incur a larger percentage of losses in the channel, particularly when the buffer size

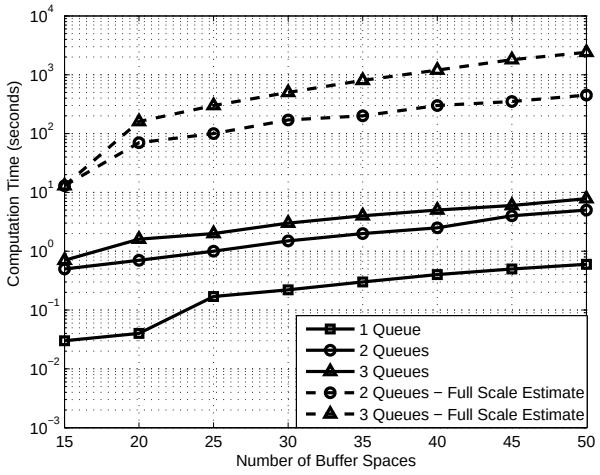


Figure 3.13: Computation Time Compared to Number of Queues

is large. However for smaller buffer sizes (*i.e.,* delay is the dominant factor), it is beneficial to target a tradeoff between types of losses.

3.4.5 Computational Complexity Discussion

System complexity can be looked at from both the first and second stage optimizations. Due to the non-generalities of non-linear programming solutions, in general it is not possible to model the complexity of the proposed NLP problem directly. We herein address this in two parts:

1) Demonstrate that the complexity of the sub optimization problem is less complex than the full-scale optimization problem, and

2) Demonstrate that the number of sub problems is less than the number of sub problems of the full-scale optimization technique.

For point 1, the full-scale optimization problem requires exploiting the joint queue state-space ($\mathcal{U}$) combined with the constraints and MAC rate state-space ($\mathcal{C}$). As the first stage of our optimization formulation does not rely on the $u \in \mathcal{U}$ states, the resulting computational complexity is not affected by the joint queue state-space size. As such, the complexity of the first-stage of our optimization is less complex than considering the full queue state-space.

The system complexity from the second stage (point 2), incurs a large complexity reduction. As we note from the previous section, the convergence of the series of LP problems occurs within several iterations. The dimension of each LP problem is $\Phi_i = |\mathcal{C}_i|(B_i + 1)$. Assuming computation time is polynomial in Φ_i, the computation time of the above is

$$T_{iter} \propto \kappa \sum_{i=1}^{K} \Phi_i \tag{3.47}$$

where κ denotes the number of iterations and $\propto$ means proportional to. Further, the computation time of the full-scale problem is

$$T_{full} \propto \prod_{i=1}^{K} \Phi_i \tag{3.48}$$

Comparing (3.47) and (3.48) its clear to see that for small κ, the computation time of (3.48) is sufficiently larger (even for $K = 2$).

In Figure 3.13 we further study the complexity by showing the computation time measured in seconds of the MAC rate assignment component for a one, two and three queue configuration averaged over 50 realizations using the iterative method. In the graph, we also provide an estimate of the computational time of both two and three queue systems assuming a full-scale optimization scheme (i.e., $\mathcal{C} \times \mathcal{U}$ as defined before) and under the assumption that full-scale optimization scales in polynomial time. We note that there is a large reduction in terms of computation time using the proposed iterative method when examining the MAC rate assignment.

3.5 Chapter Summary

In this chapter, a general $\{K \times M\}$ model is presented with a corresponding scheduler to minimize average transmission power by utilizing both channel and queue knowledge. Through a novel MAC rate assignment scheme, it is possible to decouple the problem and utilize queue state information in all queues to derive a QoS-aware scheduling policy over both static and time-varying MIMO SVD channels. The pro-

posed dynamic scheduler is shown to outperform static scheduling and is able to meet hard QoS constraints.

3.5.1 Bit-Loading Across Channels

In this chapter, bits from individual packets were loaded across multiple channels. The drawback of this approach is practical considerations from signalling. In Chapter 5, we extend this approach to packet loading which also allows use of coded modulation schemes.

3.5.2 Steady-State Assumption

The scheduling policy in this chapter is derived assuming steady-state operation of the system. While this has been shown in other works [59] that this steady state operation exists, the magnitude of the time interval to operate in steady-state has not been studied. This time interval to achieve steady state could be the study of future work.

3.5.3 Instantaneous Power Constraint

In this chapter, instantaneous power constraints were not considered as the objective was to minimize overall average power allocated. While this limits the practicality of the proposed method, the resulting optimization with a limited instantaneous power constraint may result in an infeasible solution. In the case of static channels described in this chapter, such infeasible regions exist for all time. In time-varying scenarios (such as those presented in Chapter 5), such events could be simply modeled as an outage event. This outage event is the study of future work and introduced in Chapter 8.

Chapter 4
Tracking of Sparse MIMO Channels

The application of adaptive transmission schemes (ATS) plays a vital role in the area of wireless communication systems. Recently proposed ATS [20,21] exploit knowledge of the underlying time-varying parameters of the wireless channel to adapt physical and/or medium access parameters of the channel to improve performance.

In the case of MIMO channels, knowledge of the distribution and temporal behavior of the eigenvalues of the channel matrix (and corresponding distribution and time variational properties of the capacity) are parameters of interest in the design of channel ATS. There have been a number of attempts to characterize quantities of interest such as the Level-Crossing Rate (LCR), Average Fade Duration (AFD) [60–62] and time correlation of the channel capacity [63]. Most of the publications assume rich scattering and the Jakes spectrum [64] of individual SISO channels. In addition, the covariance function of the capacity is often assumed just to follow that of the Jakes channel model [61, 63]. However, measurements [55, 56], conducted in urban environment revealed that the scattering process is often dominated by a small number of scatterers. Such channels, known as sparse channels [56] are well investigated in terms of average capacity. A geometric based MIMO model of sparse channels was presented in [65].

In this chapter, we look at the first and second order statistics, including LCR and AFD, for these sparse MIMO channels. The results presented are useful in understanding the effect of the scattering environment on cross-layer design of MIMO aware adaptive systems. Some notable results in this chapter are exploited in Chapter 5.

The remainder of the chapter is organized as follows. A general model of a sparse MIMO channel is presented in Section 4.1. First order statistics of the eigenvalues, SNR and capacity are discussed in Section 4.2 while Section 4.3 examines relevant

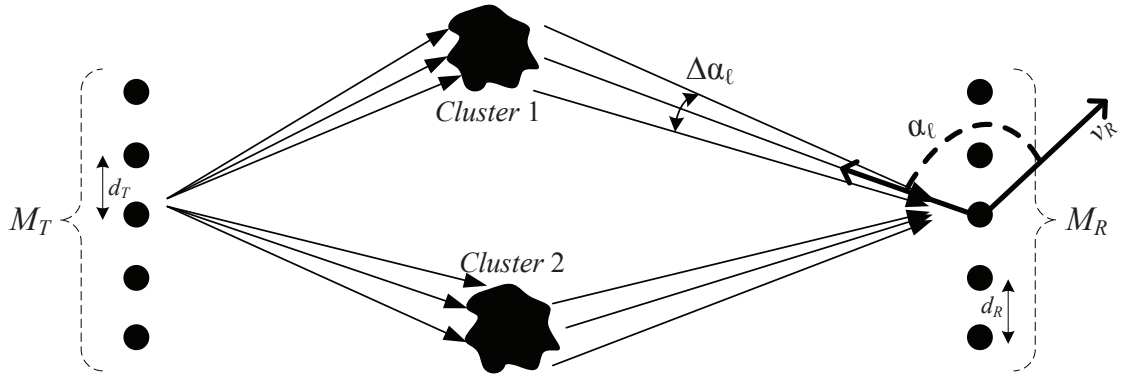


Figure 4.1: Sparse MIMO Geometric Model

time-varying parameters of these quantities. Section 4.4 details a finite state Markov model using our results. Simulation results for the statistical properties are provided in Section 4.5 and conclusions are summarized in Section 4.6. Some key results found in this chapter are also found in Dechene et al. [7].

4.1 Sparse MIMO Channel Model

Consider the scenario where the number of transmit and receive antennas, M_T and M_R respectively, is high relative to the number of clusters, where the scattering environment is dominated by a small number $L \leq \min\{M_T, M_R\}$ of angularly separable scatters (clusters) [55]. Such a system is shown in Figure 4.1. In this figure we see that the transmitted signal power reaches the receiver by scattering from two angularly separated clusters (*e.g.,* $L = 2$). The power of the scatters from each cluster is concentrated in a narrow angular width $\Delta\alpha_\ell$. The transmit and receive arrays are linear with a spacing of d_T and d_R respectively where $d_T = d_R = \lambda/2$ and where λ is the carrier wavelength. The transmit array is stationary while the receiver is moving at a speed of v_R in the direction of α_ℓ relative to direction of each cluster. The maximum Doppler frequency is f_D where

$$f_D = \frac{f_0}{c} v_R \tag{4.1}$$

and where f_0 is the carrier frequency, and c is the speed of propagation.

In this case, the MIMO channel matrix $\mathbf{H}(t)$ can be represented as [65]

$$\mathbf{H}(t) = \sum_{\ell=1}^{L} \sqrt{P_\ell}\, \mathbf{a}_\ell \mathbf{b}_\ell^\dagger \xi_\ell(t) \exp(j2\pi f_D \cos\alpha_\ell t) \tag{4.2}$$

Here α_ℓ is the angle of arrival of the signal from the ℓ^{th} scatterer, f_D is the maximum Doppler frequency, P_ℓ is the relative power of the ℓ^{th} path, where $\sum_{\ell=1}^{L} P_\ell = 1$, $\mathbf{a}_\ell$ and $\mathbf{b}_\ell$ are the receive and transmit steering vectors of the ℓ^{th} path and $\xi_\ell(t)$ is a complex Gaussian process of zero mean and a covariance function $\rho_\ell(\tau) = E\left\{\xi_\ell(t+\tau)\xi_\ell^*(t)\right\}$ given by [65]

$$\rho_\ell(\tau) = \sigma^2 \text{sinc}\left(2\Delta\alpha_\ell \cos\alpha_\ell f_D \tau\right) \exp(j2\pi f_D \cos\alpha_\ell \tau) \tag{4.3}$$

where $\Delta\alpha_\ell \ll 1$ is a narrow angular spread of ℓ^{th} cluster. The steering vectors are mutually orthogonal due to cluster separation in the angular domain, *i.e.* $\mathbf{a}_L^\dagger \mathbf{a}_k = M_R \delta_{\ell k}$, $\mathbf{b}_L^\dagger \mathbf{b}_k = M_T \delta_{k\ell}$. As the result, the matrix $\mathbf{W}(t) = \mathbf{H}(t)\mathbf{H}^\dagger(t)$ can be expanded as

$$\mathbf{W}(t) = M_T \sum_{\ell=1}^{L} \mathbf{a}_\ell \mathbf{a}_\ell^\dagger P_\ell |\xi_\ell(t)|^2 = \mathbf{A}\mathbf{P}(t)\mathbf{A}^\dagger \tag{4.4}$$

where the matrix $\mathbf{A}$, $\mathbf{A}\mathbf{A}^\dagger = M_R \mathbf{I}_{M_R}$ is composed of the steering vectors $\mathbf{a}_\ell$ and $\mathbf{P}(t) = \text{diag}\left\{P_\ell M_T |\xi_\ell(t)|^2\right\}$ is a diagonal matrix. By inspection of equation (4.4) and the fact that $L \leq \min\{M_T, M_R\}$ one can conclude that it represents eigenvalue-decomposition of the matrix $\mathbf{W}$. To compute the eigenvalues for MIMO SVD transmission, consider the following manipulations of (4.4).

Let $\mathbf{H}(t) = \mathbf{U}\mathbf{\Sigma}(t)\mathbf{V}^\dagger$ be the SVD decomposition of the channel matrix $\mathbf{H}(t)$. Here both $\mathbf{U}$ and $\mathbf{V}$ are unitary. From (4.4) we have

$$\begin{aligned} \mathbf{W}(t) = \mathbf{H}(t)\mathbf{H}(t)^\dagger &= \mathbf{U}\mathbf{\Sigma}(t)\mathbf{V}^\dagger(\mathbf{V}\mathbf{\Sigma}(t)\mathbf{U}^\dagger) & (4.5) \\ &= \mathbf{U}\mathbf{\Sigma}^2(t)\mathbf{U}^\dagger = \mathbf{A}\mathbf{P}(t)\mathbf{A}^\dagger & (4.6) \end{aligned}$$

Recalling that $\mathbf{A}\mathbf{A}^\dagger = M_R\mathbf{I}_{M_R}$, the above becomes

$$\mathbf{U}\boldsymbol{\Sigma}^2(t)\mathbf{U}^\dagger = \mathbf{X}M_R\mathbf{P}(t)\mathbf{X}^\dagger = \mathbf{X}\mathbf{Z}(t)\mathbf{X}^\dagger \tag{4.7}$$

where now $\mathbf{X}$ is unitary and $\mathbf{Z}(t) = \text{diag}\left\{P_\ell M_T M_R|\xi_\ell(t)|^2\right\}$. Clearly from the above we have

$$\boldsymbol{\Sigma}^2(t) = \text{diag}\left\{P_\ell M_T M_R|\xi_\ell(t)|^2\right\} \tag{4.8}$$

with non-zero diagonal elements of $\boldsymbol{\Sigma}^2(t)$ given as

$$\lambda_\ell^2(t) = P_\ell M_T M_R|\xi_\ell(t)|^2, \ell = 1, 2, \ldots, L \tag{4.9}$$

The above corresponds to the time-varying eigenvalues.

4.2 Statistical Properties

The above eigenvalues appear in both expressions for channel capacity, as well as they are related to the effective error rate of the individual parallel SVD channels through the SISO equivalent SNR [64]. Therefore, the statistical properties of (4.9) are useful quantities and it is important to understand both their distribution and how they evolve in time.

4.2.1 Unordered Eigenvalue and SNR Distribution

Since ξ_ℓ is a complex proper zero-mean Gaussian process with well-known [66] statistical properties, the distribution of $\eta(t) = |\xi_\ell(t)|^2$ is also well-known [66] as an exponential distribution with a scale parameter of 1 or with a PDF

$$f_{\eta_\ell}(x) = \exp(-x) \tag{4.10}$$

and by a simple transformation of variables, one can obtain the distribution of (4.9) as

$$f_{\Lambda_\ell}(x) = \frac{1}{M_R M_T P_\ell}\exp\left(-\frac{x}{M_R M_T P_\ell}\right) \tag{4.11}$$

Consequently, one can obtain equivalent SNR distribution of each subchannel. For an SVD system with equal power allocation, and assuming a reference SNR level of γ_0, one can relate this to the eigenvalues as [67]

$$\gamma_\ell = \frac{\gamma_0}{M_T}\lambda_\ell \tag{4.12}$$

with a corresponding per subchannel SNR density function at a reference SNR γ_0 given through a trivial transformation of variables as

$$f_{\Gamma_\ell}(\gamma_\ell) = \frac{1}{\gamma_0 M_R P_\ell}\exp\left(-\frac{\gamma_\ell}{\gamma_0 M_R P_\ell}\right) \tag{4.13}$$

The above SNR density can be used to determine performance parameters (such as bit-error rate) of the individual channels.

4.2.2 Ordered Eigenvalue and SNR Distribution

In addition to the individual distributions, the ordered eigenvalue distribution is often a quantity of interest. We denote the ordered eigenvalues as

$$\lambda'_L \geq \lambda'_{L-1} \geq \cdots \geq \lambda'_1 \geq 0 \tag{4.14}$$

With equal power arriving from all clusters (*i.e.* $P_\ell = 1/L, \ell = 1, 2, \ldots, L$), the k^{th} eigenvalue can be found as [64]

$$f_{\lambda'_k}(x) = k\frac{L!}{k!(L-k)!}F_X^{k-1}(x)[1 - F_X(x)]^{L-k}f_X(x) \tag{4.15}$$

where

$$f_X(x) = \frac{L}{M_R M_T}\exp\left(-\frac{xL}{M_R M_T}\right) \tag{4.16}$$

and

$$F_X(x) = \int_0^x f_X(\bar{x})d\bar{x} = 1 - \exp\left(-\frac{xL}{M_R M_T}\right) \tag{4.17}$$

and similar expressions can easily be found for SNR.

In the general case when eigenvalues/SNR are not identically distributed, the ordered distribution can be obtained by applying the Bapat-Beg Theorem [68]. In general however, such expressions must be solved numerically and are not the focus of this work.

4.2.3 Channel Capacity

It is well known [67] that the instantaneous capacity of an open loop MIMO system is given as

$$C(t) = \ln \det \left[\mathbf{I}_{M_R} + \frac{\gamma_0}{M_T} \mathbf{W}(t) \right] \tag{4.18}$$

with all quantities given as before. Furthermore, (4.18) can also be expressed a summation of terms containing the eigenvalues of the MIMO channel as

$$C(t) \;=\; \sum_{\ell=1}^{L} \ln \left[1 + \gamma_0 M_R P_\ell \eta_\ell(t) \right] \tag{4.19}$$

As such, statistics of $C(t)$ could be studied in terms of statistics of $\eta_\ell(t)$.

4.2.3.1 Capacity Distribution

In the above, (4.19) shows that the contribution to capacity for sparse MIMO channels is the summation of capacity contributions by individual scatters. Using this principle, we derive the capacity distribution as follows.

As the first step, let us consider the capacity contribution by any single scatter ℓ (*i.e.*, $C_\ell(t)$). It is given as

$$C_\ell(t) = \ln[1 + \bar{\gamma_\ell} \eta_\ell(t)] \tag{4.20}$$

where $\bar{\gamma_\ell} = \gamma_0 M_R P_\ell$ and the distribution of $\eta_\ell(t)$ is exponential as in (4.10). Let $F_{C,\ell}(c)$ denote the distribution function of the capacity contribution of the ℓ^{th} cluster. The distribution $F_{C,\ell}(c)$ can be found through a transformation of random variables

as follows.

$$\begin{aligned} F_{C,\ell}(c) &= Pr[C \leq c] && (4.21)\\ &= Pr\left[\ln(1+\bar{\gamma_\ell}\eta_\ell) \leq c\right] && (4.22)\\ &= Pr\left[\eta_\ell \leq \frac{\exp(c)-1}{\bar{\gamma_\ell}}\right] && (4.23)\\ &= F_{\eta_\ell}\left(\frac{\exp(c)-1}{\bar{\gamma_\ell}}\right) && (4.24)\\ &= 1-\exp\left(-\frac{\exp(c)-1}{\bar{\gamma_\ell}}\right) && (4.25) \end{aligned}$$

where the PDF is then easily found as

$$f_{C,\ell}(c) = \frac{\partial}{\partial c}\left[F_{C,\ell}(c)\right] = \frac{1}{\bar{\gamma_\ell}}\exp\left(-\frac{\exp(c)-1}{\bar{\gamma_\ell}}+c\right) \quad (4.26)$$

Finally since the contribution by each cluster is independent, the overall distribution is

$$f_C(x) = f_{C,1}(c_1) * f_{C,2}(c_2) * \cdots * f_{C,L}(c_L) \quad (4.27)$$

where $*$ denotes the convolution operation.

4.2.3.2 Exact Mean and Variance of Capacity

While direct computation of the above provides information about the capacity distribution, it is often [69] more efficient to approximate the above as a Gaussian distribution, which can be completely described through its mean and variance.

From (4.19) we can obtain the exact expression for the mean capacity, $\mu_{\bar{C}}$, as follows

$$\mu_{\bar{C}} = E[C(t)] = E\left[\sum_{\ell=1}^{L}\ln(1+\bar{\gamma_\ell}\eta_\ell(t))\right] \quad (4.28)$$

$$= \sum_{\ell=1}^{L} E\left[\ln(1+\bar{\gamma_\ell}\eta_\ell(t))\right] \tag{4.29}$$

$$= \sum_{\ell=1}^{L} \int_0^\infty \ln(1+\bar{\gamma_\ell}x)\exp(-x)dx \tag{4.30}$$

In a similar manner, the second moment of capacity is given as

$$E[C^2(t)] = \sum_{\ell=1}^{L} \int_0^\infty \ln^2(1+\bar{\gamma_\ell}x)\exp(-x)dx \tag{4.31}$$

Therefore the variance σ_C^2 is given as

$$\sigma_C^2 = E[C^2(t)] - E^2[C(t)] \tag{4.32}$$

To evaluate both (4.30) and (4.31), let us consider an integral of the form

$$I_m = \int_0^\infty \ln^m(1+\gamma) f_\gamma(\gamma)d\gamma \tag{4.33}$$

where $f_{\bar{\gamma_\ell}}(\gamma)$ is the PDF of the random variable γ and relating the above to the integrals in (4.30) and (4.31) is $f_{\bar{\gamma_\ell}}(\gamma) = {\bar{\gamma_\ell}}^{-1}\exp\left(-\gamma/\bar{\gamma_\ell}\right)$

Exact evaluation of the integral (4.33) could be based on the following identity

$$\ln^m(1+\gamma) = \lim_{\nu\to 0}\frac{d^m}{d\nu^m}(1+\gamma)^\nu \tag{4.34}$$

Therefore, evaluation of I_m can be reduced to evaluation of

$$J = \int_0^\infty (1+\gamma)^\nu f_{\bar{\gamma_\ell}}(\gamma)d\gamma \tag{4.35}$$

followed by differentiation and taking the limit

$$I_m = \lim_{\nu\to 0}\frac{d^m}{d\nu^m}J \tag{4.36}$$

In particular, given the exponential distribution above for the capacity contribution from cluster ℓ, one can obtain the following

$$J^{(\ell)} = \exp\left(\frac{1}{\bar{\gamma}_\ell}\right) \bar{\gamma_\ell}^{\nu} \Gamma\left(1+\nu, \frac{1}{\bar{\gamma}_\ell}\right) \tag{4.37}$$

where $\Gamma(x, a)$ is the incomplete Gamma function.

For $m = 1$, we have

$$I_1^{(\ell)} = \lim_{\nu \to 0} \frac{d}{d\nu} \exp\left(\frac{1}{\bar{\gamma}_\ell}\right) \bar{\gamma_\ell}^{\nu} \Gamma\left(1+\nu, \frac{1}{\bar{\gamma}_\ell}\right) = \exp\left(\frac{1}{\bar{\gamma}_\ell}\right) \mathrm{E}_1\left(\frac{1}{\bar{\gamma}_\ell}\right) \tag{4.38}$$

where $\mathrm{E}_1(\cdot)$ is the exponential integral [70] defined as

$$\mathrm{E}_1(x) \equiv \int_x^\infty \frac{\exp(-u)du}{u} \tag{4.39}$$

Routines for numerical evaluation of $E_1(x)$ is available in most mathematical software packages. Using (4.30) and (4.39) one can obtain an expression for the mean of capacity as

$$\bar{\mu_C} = \sum_{\ell=1}^{L} I_1^{(\ell)} = \sum_{\ell=1}^{L} \exp\left(\frac{1}{\bar{\gamma}_\ell}\right) \mathrm{E}_1\left(\frac{1}{\bar{\gamma}_\ell}\right) \tag{4.40}$$

Similarly, one can obtain the second moment of capacity from (4.31) as

$$E[C^2(t)] = \sum_{\ell=1}^{L} I_2^{(\ell)} = \lim_{\nu \to 0} \frac{d^2}{d\nu^2} J^{(\ell)} = \sum_{\ell=1}^{L} 2 \exp\left(\frac{1}{\bar{\gamma}_\ell}\right) G_{23}^{30}\left(\begin{array}{c} 1,1 \\ 0,0,0 \end{array} \middle| \frac{1}{\bar{\gamma}_\ell} \right) \tag{4.41}$$

where $G_{pq}^{mn}\left(\begin{array}{c} a_1, \ldots, a_p \\ b_1, \ldots, b_q \end{array} \middle| z \right)$ is the Meijer-G function [70], which is readily available in many numerical packages. Using the above, the variance of capacity is simply as usual

$$\sigma_C^2 = \sum_{\ell=1}^{L} \left[2 \exp\left(\frac{1}{\bar{\gamma}_\ell}\right) G_{23}^{30}\left(\begin{array}{c} 1,1 \\ 0,0,0 \end{array} \middle| \frac{1}{\bar{\gamma}_\ell} \right) - \left(\exp\left(\frac{1}{\bar{\gamma}_\ell}\right) \mathrm{E}_1\left(\frac{1}{\bar{\gamma}_\ell}\right) \right)^2 \right] \tag{4.42}$$

The above variance requires numerical computation. In the next sections, we examine different approaches for tractable computation of both the mean and variance of capacity.

4.2.3.3 Approximate Mean and Variance - Moment-Based Approach

Alternatively, one can directly apply the moment-based method in [71] to obtain a tractable approximation for the exact mean and variance of capacity found above. In this case, one can express the Taylor series about the average SNR value $\bar{\gamma_\ell}$ of the quantity in (4.33) as

$$\ln(1+\gamma) = \ln(1+\bar{\gamma_\ell}) + \sum_{\omega=1}^{\infty} \frac{(-1)^{\omega-1}(\gamma-\bar{\gamma_\ell})^\omega}{\omega(1+\bar{\gamma_\ell})^\omega} \tag{4.43}$$

$$\approx \ln(1+\bar{\gamma_\ell}) + \frac{\gamma-\bar{\gamma_\ell}}{1+\bar{\gamma_\ell}} - \frac{(\gamma-\bar{\gamma_\ell})^2}{2(1+\bar{\gamma_\ell})^2} + \mathcal{O}\left[(\gamma-\bar{\gamma_\ell})^2\right] \tag{4.44}$$

The mean is given as before in (4.30) and is

$$\bar{\mu_C} = \sum_{\ell=1}^{L} \int_0^\infty \ln(1+\bar{\gamma_\ell}x)\exp(-x)dx \tag{4.45}$$

$$= \sum_{\ell=1}^{L} \int_0^\infty \ln(1+\gamma) f_{\gamma_\ell}(\gamma) \tag{4.46}$$

$$= \sum_{\ell=1}^{L} \int_0^\infty \left(\ln(1+\bar{\gamma_\ell}) + \sum_{\omega=1}^{\infty} \frac{(-1)^{\omega-1}(\gamma-\bar{\gamma_\ell})^\omega}{\omega(1+\bar{\gamma_\ell})^\omega}\right)\exp(-\gamma)d\gamma \tag{4.47}$$

$$= \ln(1+\bar{\gamma_\ell}) + \sum_{\omega=1}^{\infty} \frac{(-1)^{\omega-1}}{\omega(1+\bar{\gamma_\ell})^\omega}\mu_\omega \tag{4.48}$$

where μ_ω is the ω^{th} central moment of γ. Taking an approximation to $\omega = 2$ we have

$$\bar{\mu_C} \approx \sum_{\ell=1}^{L} \ln(1+\bar{\gamma_\ell}) - \frac{\bar{\gamma_\ell}^2}{2(1+\bar{\gamma_\ell})^2} + \mathcal{O}\left[\frac{\mu_3}{3(1+\bar{\gamma_\ell})^3}\right] \tag{4.49}$$

as by definition $\mu_1 = 0$. Similarly, this approach can be used for the second moment

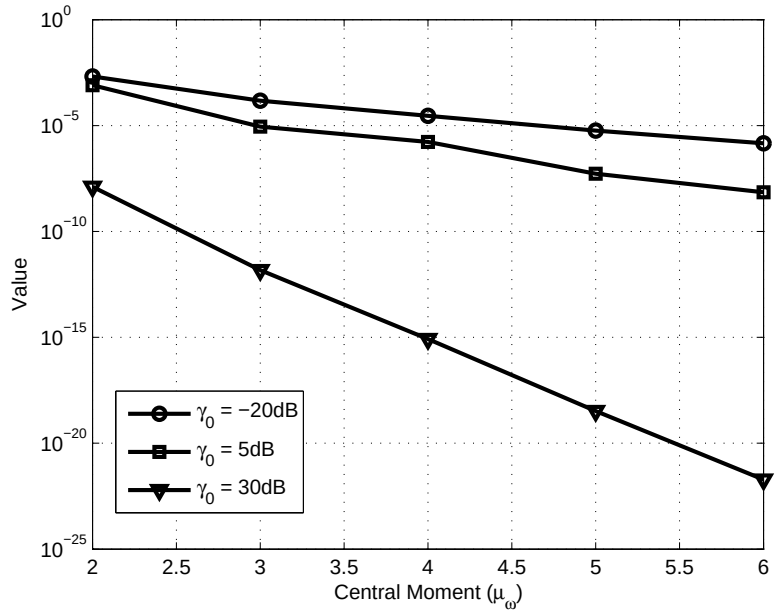


Figure 4.2: Impact of Higher Order Moments on Expansion Coefficients

as

$$\sigma_C^2 + \bar{\mu_C}^2 = \sum_{\ell=1}^{L} \int_0^{\infty} \left(\ln(1+\bar{\gamma_\ell}) + \sum_{\omega=1}^{\infty} \frac{(-1)^{\omega-1}(\gamma-\bar{\gamma_\ell})^\omega}{\omega(1+\bar{\gamma_\ell})^\omega} \right)^2 f_{\bar{\gamma_\ell}}(\gamma) d\gamma \quad (4.50)$$

where after reducing the above to $\omega = 2$ and solving for the variance, we obtain

$$\sigma_C^2 \approx \sum_{\ell=1}^{L} \frac{\bar{\gamma_\ell}^2}{(1+\bar{\gamma_\ell})^2} - \frac{\bar{\gamma_\ell}^4}{4(1+\bar{\gamma_\ell})^4} \quad (4.51)$$

where model validations are found in Section 4.5.

It is important to study the factor $\frac{\mu_\omega}{\omega(1+\gamma_\ell)^\omega}$ as in general central moments may increase without limit and limiting the number of terms may dramatically affect the accuracy of the approximation, particularly for moderate to high SNR. The value of the above factor for increasing ω is shown in Figure 4.2 for $L = 1, M_R = 8$. Due to the difficulty in obtaining an analytical solution, these values are computed using numerical techniques. In this Figure, we observe that even under high SNR conditions

(30dB), the factor decays for increasing ω as a result of the denominator exponent. As a result, we suggest that solving to $\omega = 2$ may provide a good approximation to the Taylor expansion in (4.33).

4.2.3.4 Approximate Mean and Variance - Low and High SNR Asymptotics

A different approach to the approximation for the low and high SNR region is to use the following approximations

$$\log(1+x) \approx \begin{cases} x, & x \ll 1 \text{ (Low SNR)} \\ \log(x), & x \gg 1 \text{ (High SNR)} \end{cases} \tag{4.52}$$

Using the above, the mean in the low SNR region reduces to

$$\bar{\mu_C}^{(Low)} = \sum_{\ell=1}^{L} \int_0^{\infty} \gamma f_{\bar{\gamma_\ell}}(\gamma) d\gamma = \sum_{\ell=1}^{L} \bar{\gamma_\ell} \tag{4.53}$$

and similarly one can obtain the variance as

$$\sigma_C^{2\,(Low)} = \sum_{\ell=1}^{L} \gamma_\ell^2 \tag{4.54}$$

In the high SNR region, one can use a similar approach to the exact computation with $\ln(1+\gamma) \approx \ln(\gamma)$. In this case, we obtain

$$J^{(\ell)} = \bar{\gamma_\ell}^{v} \Gamma(1+v) \tag{4.55}$$

where $\Gamma(x)$ is the Gamma function. Following arguments as before we have

$$\bar{\mu_C}^{(High)} = \sum_{\ell=1}^{L} \left(\ln(\gamma_\ell) - \bar{\gamma_0}\right) \tag{4.56}$$

where $\bar{\gamma_0} = 0.577215665...$ is the Euler-Mascheroni constant [70].

The high SNR variance is found by taking the asymptotic expansion of

$$\lim_{v\to 0}\frac{d^2}{dv^2}\left(J^{(\ell)}\right)-\left(I_1^{(\ell)}\right)^2$$

To the second order, this is given as

$$\sigma_C^{2\ (High)}=\sum_{\ell=1}^{L}\frac{2[\bar{\gamma_0}-\ln(\bar{\gamma_\ell})-(\ln(\bar{\gamma_\ell})-\bar{\gamma_0})^2]}{\bar{\gamma_\ell}}+\frac{\pi^2}{6} \tag{4.57}$$

From the above its easy to see that the variance converges as SNR goes to infinity

$$\sigma_C^2=\lim_{\gamma_\ell\to\infty,\forall\ell}=\sum_{\ell=1}^{L}\frac{\pi^2}{6}=\frac{L\pi^2}{6} \tag{4.58}$$

This is consistent with the results obtained in [72].

4.3 Time-Varying Statistics

In addition to time-independent statistics, it is often useful in the design of adaptive transmission schemes to utilize time-varying parameters [20, 21] including the level-crossing rate (LCR) and the average fade duration (AFD). These quantities describe useful characteristics about the how quickly the channel changes in time and how these changes affect the reliability of the channel.

4.3.1 Unordered Eigenvalues

It follows from (4.3) that for any unordered eigenvalue, the correlation function of the envelope of ξ_ℓ is given by

$$\rho_{env,\ell}(\tau)=\sigma^2\text{sinc}\left(2\Delta\alpha_\ell\cos\alpha_\ell f_D\tau\right) \tag{4.59}$$

It is well known [66] that the LCR of a single eigenvalue with a exponential density function given in (4.11) is exactly

$$N_{\Lambda,\ell}(\lambda) = 2\sqrt{\frac{\lambda}{M_R M_T P_\ell}} \exp\left(\frac{-\lambda}{M_R M_T P_\ell}\right) \sqrt{\frac{\dot{\sigma}_\ell^2}{2\pi}} \tag{4.60}$$

where $\dot{\sigma}_\ell^2$ is the variance of the derivative process. For a Gaussian process, this is known to be

$$\begin{aligned} \dot{\sigma}_\ell^2 &= -\frac{\partial^2}{\partial \tau^2}\{\rho_{env,\ell}(\tau)\}\Big|_{\tau=0} & (4.61)\\ &= \frac{\sigma^2}{3}(2\pi\Delta\alpha_\ell \cos\alpha_\ell f_D)^2 & (4.62) \end{aligned}$$

with corresponding AFD as

$$AFD_{\Lambda,\ell}(\lambda) = \frac{F_{\Lambda,\ell}(\lambda)}{N_{\Lambda,\ell}(\lambda)} \tag{4.63}$$

where $F_{\Lambda,\ell}(\lambda)$ is the CDF of the exponential distribution given in (4.11) and equal to

$$F_{\Lambda,\ell}(x) = 1 - \exp\left(-\frac{x}{M_R M_T P_\ell}\right) \tag{4.64}$$

4.3.2 Capacity

4.3.2.1 Single Cluster Capacity LCR and AFD

An exact expression for LCR and AFD of channel capacity for a single cluster environment can be found as follows. First the channel capacity of a single cluster is given as before

$$C_\ell(t) = \ln(1 + \bar{\gamma_\ell}\eta_\ell(t)) \tag{4.65}$$

As with (4.60), the LCR of the $\bar{\gamma_\ell}\eta_\ell(t)$ process is also exponentially distributed and given as

$$N_{\bar{\gamma_\ell}\eta_\ell}(r) = 2\sqrt{\frac{r}{\bar{\gamma_\ell}}} \exp\left(\frac{-r}{\bar{\gamma_\ell}}\right) \sqrt{\frac{\dot{\sigma}_\ell^2}{2\pi}} \tag{4.66}$$

where all quantities are as before. Due to the monotonically increasing and non-negative nature of (4.65), there exists a one-to-one mapping of $C_\ell(t)$ to $\eta_\ell(t)$. Further, the LCR at a threshold c, where

$$c = \ln(1 + r) \tag{4.67}$$

is equal to LCR at a threshold r given in (4.66). Since $r = \exp(c) - 1$, the LCR of capacity of a single cluster is given as

$$N_{C_\ell}(c) = N_{\bar{\gamma_\ell}\eta_\ell}(\exp(c) - 1) = \sqrt{\frac{2\beta_\ell(\exp(c) - 1)}{\pi\bar{\gamma_\ell}}} \exp\left(-\frac{\exp(c) - 1}{\bar{\gamma_\ell}}\right) \tag{4.68}$$

and the AFD is given as usual

$$AFD_{C_\ell}(c) = \frac{F_{C_\ell}(c)}{N_{C_\ell}(c)} \tag{4.69}$$

where $F_{C_\ell}(c)$ is the cumulative density function of (4.26).

4.3.2.2 Approximation of Multi-Cluster Capacity LCR and AFD

It was noted in Section 4.2 that the contribution of individual clusters to the capacity is independent. However, this is not the case in the computation of LCR and AFD, therefore requiring direct use of the general formula Rice formula [66] which requires knowledge of the joint PDF of the capacity and its derivative process (which is not usually known [61,73]). Since a derivative of a stationary process is often a symmetric function uncorrelated with the original process, it is usually the case that LCR can be approximated as

$$N_C(c) = p_c(c)\sqrt{\frac{\sigma_{\dot{c}}^2}{2\pi}} \tag{4.70}$$

where $\sigma_{\dot{c}}^2$ is the variance of the capacity derivative process.

Assuming the validity of (4.70) the approximation for $\sigma_{\dot{c}}^2$ is given by the following lemma:

Lemma 1. *The capacity derivative process $\dot{C}$ is approximately a zero-mean Gaussian random variable with an approximate variance of*

$$\sigma_{\dot{c}}^2 \approx 4\exp(-1)\pi\sigma_c^2 \sum_{\ell=1}^{L} P_\ell \dot{\sigma}_\ell^2$$

Justification: Firstly, as before we note that due to the non-negative nature of the underlying Rayleigh process $|\xi_\ell(t)|$, there exists a one to one mapping between $C_\ell(t)$ and $|\xi_\ell(t)|$ such that at any instant of time, $C_\ell(t) = \log(1+\bar{\gamma_\ell}|\xi_\ell(t)|^2)$. Furthermore, the non-negative property of $|\xi_\ell(t)|$ implies that the average number of times the $|\xi_\ell(t)|$ process crosses a threshold r is equal to the average number of times the $C_\ell(t)$ process crosses a threshold c where $c = \log(1+\gamma_0 M_R P_\ell |r|^2)$.

Under this observation, we suggest that the maximum value of the LCR of the underlying Rayleigh process is the same as the capacity process. Under our Gaussian capacity approximation, where the capacity variance is σ_c^2, we obtain a maximum distribution value of $1/\sqrt{2\pi\sigma_c^2}$ where the LCR peaks. Using this feature matching method, the ratio of the underlying Rayleigh process (where $\sigma^2 = 0.5$) maximum to the Gaussian capacity is:

$$\kappa = \sqrt{2}\exp(-\frac{1}{2})\sqrt{2\pi\sigma_c^2}$$

Finally, one can obtain the approximate process variance as a weighted combination of the variances of each individual contribution of capacity scaled by the distribution ratio above

$$\begin{aligned}\sigma_{\dot{c}}^2 &\approx \sum_{\ell=1}^{L} \kappa^2 P_\ell \dot{\sigma}_\ell^2 \\ &\approx 4\exp(-1)\pi\sigma_c^2 \sum_{\ell=1}^{L} P_\ell \dot{\sigma}_\ell^2\end{aligned}$$

where all quantities have been previously described.

Using the LCR in (4.70), the AFD is simply

$$AFD_C(c) = \frac{F_C(c)}{N_C(c)} \tag{4.71}$$

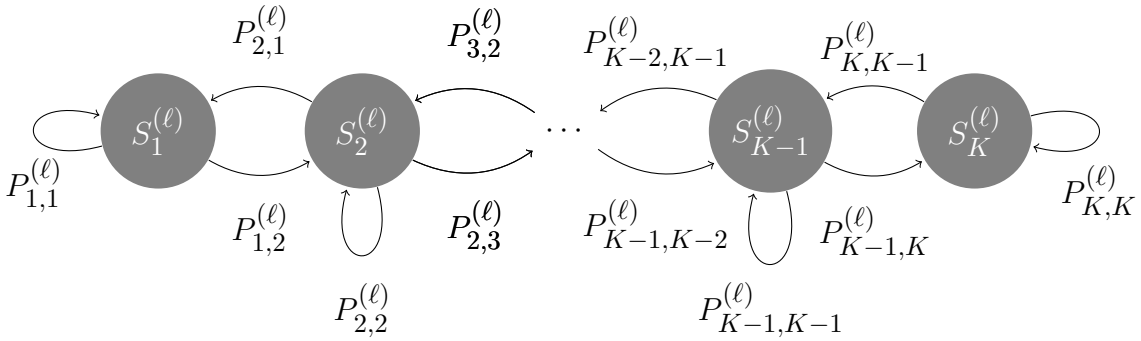


Figure 4.3: Eigenvalue Evolution for the ℓ^{th} eigenvalue

Here $F_C(c)$ is a Gaussian CDF with mean and variance given in (4.40) and (4.42) respectively.

4.4 Finite State Channel Model

Alone, quantities such as the LCR and AFD of a process can be of little use. Combining this with use of finite-state Markov chains (FSMCs) [19] forms the well-known Wang-Moyaeri model, and provides of mechanism of quantifying the dynamics in a form useful for cross-layer scheduling techniques [20, 21].

4.4.1 Eigenvalue Based

The unordered channel eigenvalues can be described as an L-dimensional FSMC which models the evolution of the eigenvalues as a function of time. By partitioning each eigenvalue process into a finite level of states, with corresponding probabilities of transmission between states, we can model the evolution of the entire process. Such a model is vital in the design of adaptive systems such as those which employ SVD eigenmode transmission as the error performance of the system relies on the instantaneous eigenvalues.

The design of the L-FSMC model is as follows. For each of the L eigenvalues, partition the process into K finite states denoted $\mathcal{S}_\ell \in \{S_k^{(\ell)}\}, k = 1, 2, \ldots, K$ where

each state is bounded by $[\varphi_{\ell,i}, \varphi_{\ell,i+1})$ with $\varphi_{\ell,1} = 0$ and $\varphi_{\ell,K+1} = \infty$ for each ℓ^{th} eigenvalue. There are numerous partitioning methods discussed in the literature [19, 20], however for the sake of brevity we do not discuss those here and assume that we have chosen these thresholds. Overall we obtain the entire L-FSMC represented by K^L independent states (or by $\mathcal{S} = \mathcal{S}_1 \times \cdots \times \mathcal{S}_L$).

The transition probabilities between states can be computed in a manner such as that in [19]. We denote the frame of interest having a duration T_f. We also assume during a frame of interest, each eigenvalue can evolve to only a neighbouring state. In this way, the transition probabilities for each of the L individual eigenvalues can be given as:

$$P_{i,i+1}^{(\ell)} \approx \frac{N_{\Lambda,\ell}(\varphi_{\ell,i+1})T_f}{F_{\Lambda,\ell}(\varphi_{\ell,i+1}) - F_{\Lambda,\ell}a(\varphi_{\ell,i})} \quad i = 1, \ldots, K-1 \tag{4.72}$$

$$P_{i,i-1}^{(\ell)} \approx \frac{N_{\Lambda,\ell}(\varphi_{\ell,i})T_f}{F_{\Lambda,\ell}(\varphi_{\ell,i+1}) - F_{\Lambda,\ell}(\varphi_{\ell,i})} \quad i = 2, \ldots, K \tag{4.73}$$

$$P_{i,i}^{(\ell)} \approx 1 - P_{i,i-1}^{(\ell)} - P_{i,i+1}^{(\ell)} \quad i = 2, \ldots, K-1 \tag{4.74}$$

$$P_{1,1}^{(\ell)} \approx 1 - P_{1,2}^{(\ell)} \quad i = 1 \tag{4.75}$$

$$P_{K,K}^{(\ell)} \approx 1 - P_{K,K-1}^{(\ell)} \quad i = K \tag{4.76}$$

where the evolution for each eigenvalue is shown in Figure 4.3 and the overall eigenvalue Markov chain is comprised of L of these independent eigenvalue chains.

4.4.2 Capacity Based

The study of channel capacity is often of equal interest when analyzing the performance of a given channel. This is no different when quantifying how it changes in time and has been used in a number of ATSs [74].

In this case, the channel capacity can be modeled as a one-dimensional FSMC. The FSMC is comprised of K finite states with each bounded by $[\varphi_i, \varphi_{i+1})$ and with $\varphi_1 = 0$ and $\varphi_{K+1} = \infty$. The corresponding state transition probabilities are as before

and given as

$$P_{i,i+1} \approx \frac{N_C(\varphi_{i+1})T_f}{F_C(\varphi_{i+1}) - F_C(\varphi_i)} \quad i = 1, \ldots, K-1 \tag{4.77}$$

$$P_{i,i-1} \approx \frac{N_C(\varphi_i)T_f}{F_C(\varphi_{i+1}) - F_C(\varphi_i)} \quad i = 2, \ldots, K \tag{4.78}$$

$$P_{i,i} \approx 1 - P_{i,i-1} - P_{i,i+1} \quad i = 2, \ldots, K-1 \tag{4.79}$$

$$P_{1,1} \approx 1 - P_{1,2} \quad i = 1 \tag{4.80}$$

$$P_{K,K} \approx 1 - P_{K,K-1} \quad i = K \tag{4.81}$$

4.5 Numerical Validation

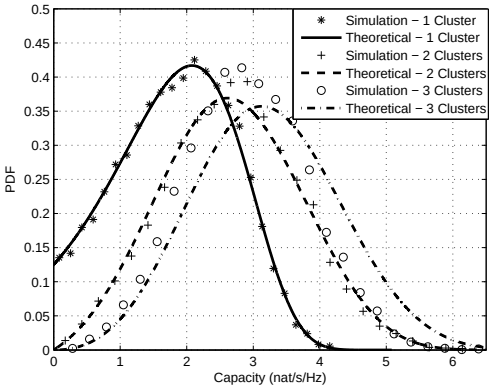
In order to evaluate the accuracy of our model and approximations, simulation analysis is performed. All default parameters are listed in Table 4.1.

4.5.1 Time-Invariant Statistics

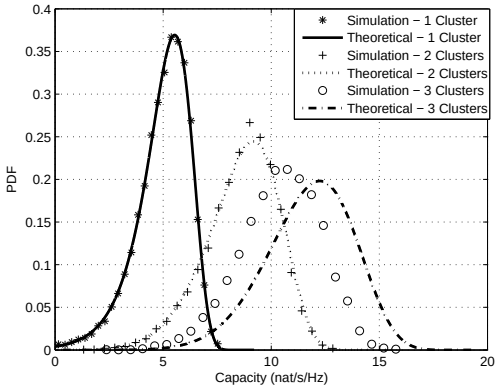
In Figure 4.4 we demonstrate the simulated and analytical capacity distribution for various antenna configurations. We observe that for the most part the analytical distribution closely matches the simulated results. For the case of the 8×8 system with 3 clusters we observe this is not exactly the case. In this case, the number of clusters relative to the number of antennas is increasing resulting in incomplete separability of the individual clusters. Further, for the 2 and 3 cluster case, we apply

Table 4.1: Simulation Parameters

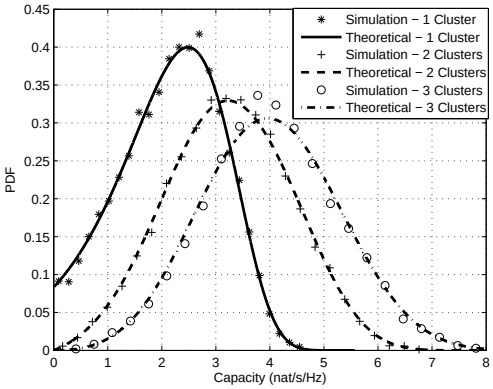
Parameter	Value	Parameter	Value
Number of Samples	20000	γ_0	0dB
F_s	$1250Hz$	v_R	$60km/hr$
Antenna Spacing	$\lambda/2$	f_C	$2.4GHz$
$\Delta\alpha_\ell,\ \forall\ell$	5°	$P_\ell,\ \forall\ell$	$1/L$
α_1	0°	α_2	30°
α_3	60°		



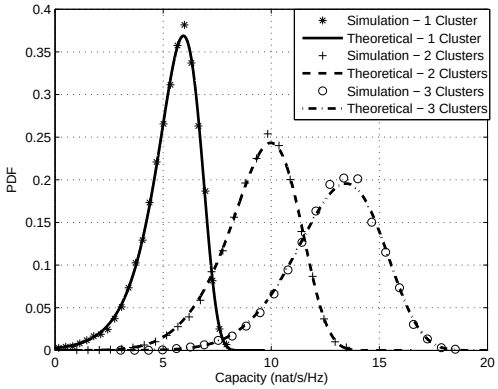
(a) 8x8 MIMO - 0dB



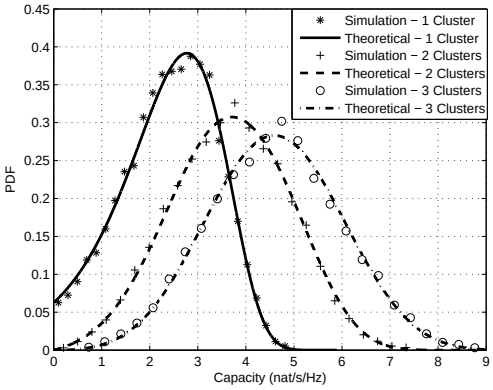
(b) 8x8 MIMO - 15dB



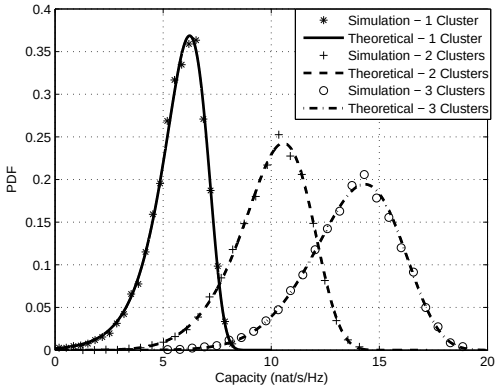
(c) 12x12 MIMO - 0dB



(d) 12x12 MIMO - 15dB

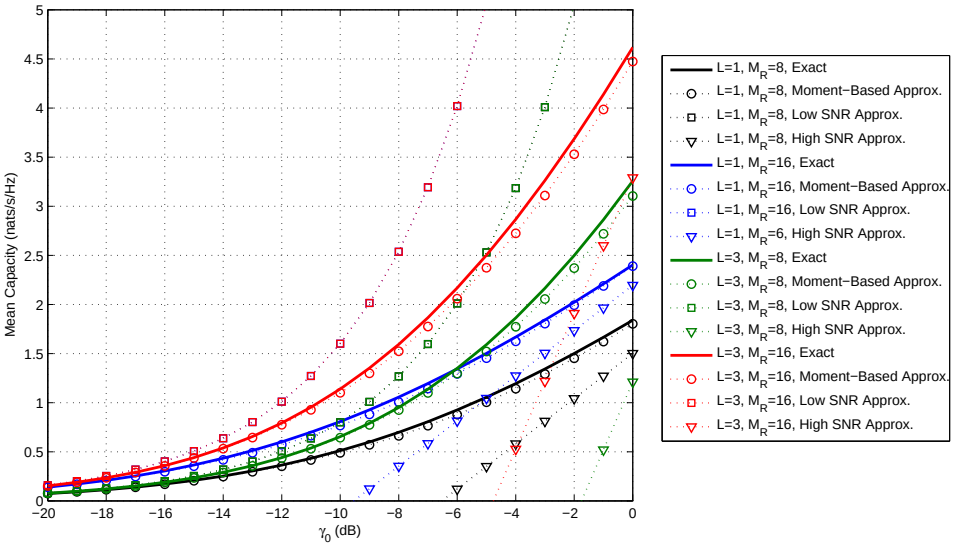


(e) 16x16 MIMO - 0dB

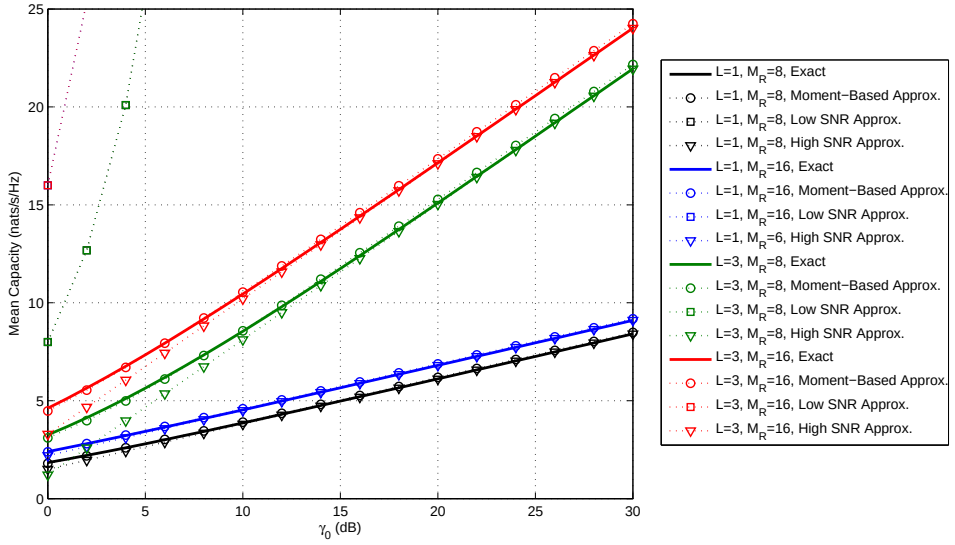


(f) 16x16 MIMO - 15dB

Figure 4.4: Capacity Distribution for Various Antenna and SNR Values

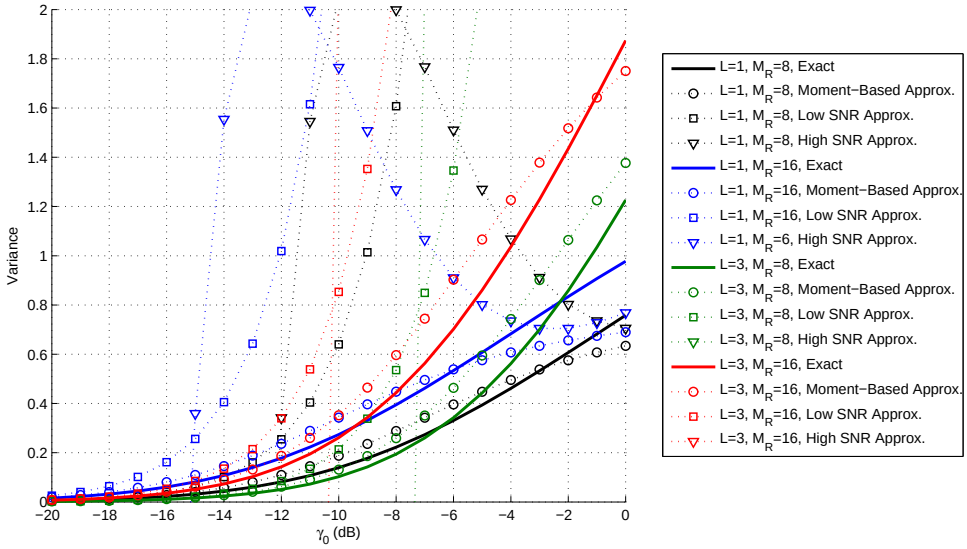


(a) Low SNR Region

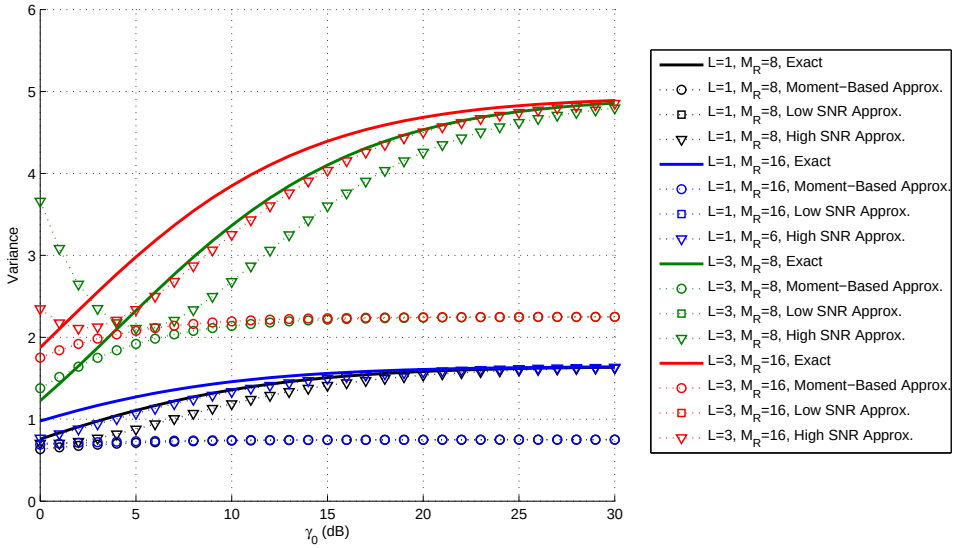


(b) High SNR Region

Figure 4.5: Mean Channel Capacity as a Function of SNR



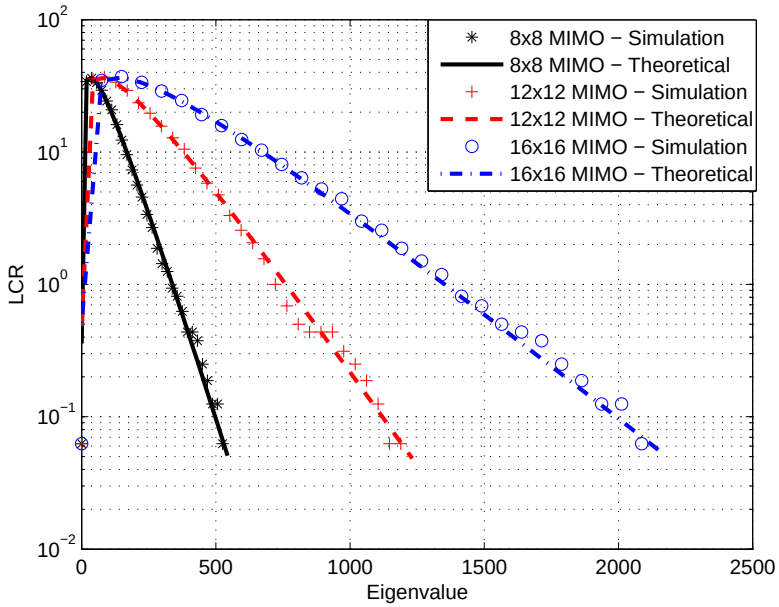
(a) Low SNR Region



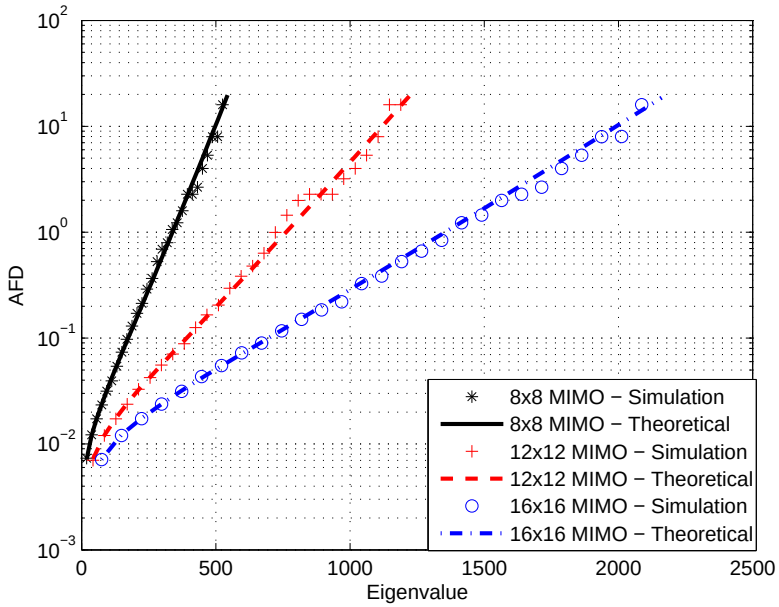
(b) High SNR Region

Figure 4.6: Variance of Channel Capacity as a Function of SNR

the KS-TEST with a significance level of 0.05 to verify the accuracy of the Gaussian approximation. The KS-TEST fails to reject this approximation under all scenarios, except for the case of 8×8 system with 3 clusters. The rationale behind this is as



(a) LCR



(b) AFD

Figure 4.7: Single Eigenvalue Level-Crossing Rate and Average Fade Duration

before.

The capacity exact and approximate mean is shown in Figure 4.5 for low and

high SNR respectively. In the low SNR region the low SNR approximation only holds close for less than about -14dB. The moment-based method closely follows the mean from low to high SNR. In the high SNR region, the high SNR approximation very closely approximates the mean capacity for 15dB and greater.

The variance is shown in Figure 4.6. For the case of the variance, the approximations are not as close. The low SNR approximation is only valid for very low SNRs (less than 20dB). The moment-based method holds tightly until about -10dB and reasonably close until about 0dB. The high SNR approximation approaches the exact around 15dB for a single cluster, and as high as 30dB for the 3 cluster scenario.

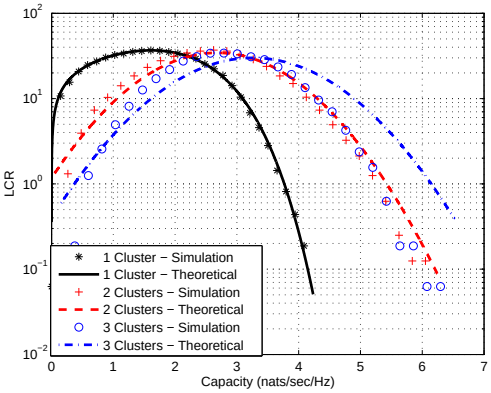
4.5.2 Time-Varying Statistics

Figures 4.7a and 4.7b shows the LCR and AFD for single eigenvalues respectively for various antenna configurations. The analytical results in this case exactly follow the simulated performance.

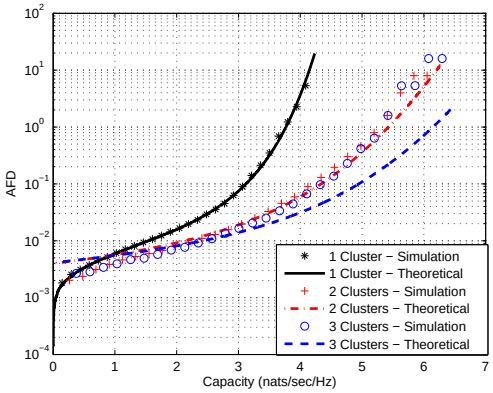
Further, figures 4.8a-4.8f show the LCR and AFD for the channel capacity for several antenna configurations. Trends of the LCR and AFD (exact and approximate) are as expected. The single cluster contribution is exact (since this is simply an extension of the eigenvalue LCR/AFD findings above). The multi-cluster results closely approximates the simulated performance justifying its use. The only scenario where there is dramatic disagreement between analytical expressions and simulated results is for $L = 3$ for 8×8 MIMO. Here we observe similar phenomenon as the capacity distribution (the independence assumption no longer holds due to clusters not being completely separable).

4.6 Chapter Summary

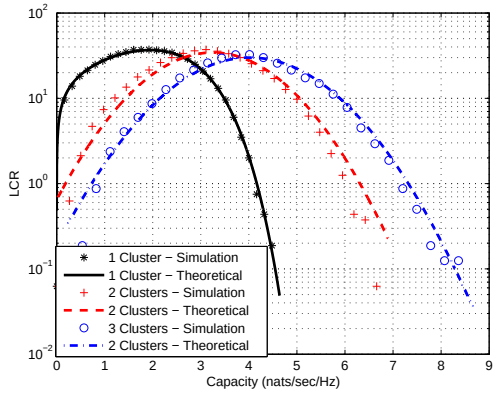
In this chapter we derived expressions for the first and second order statistics of sparse MIMO channels. We provide exact expressions for the LCR and AFD of the unordered channel eigenvalues and capacity contributed by a single narrow cluster.



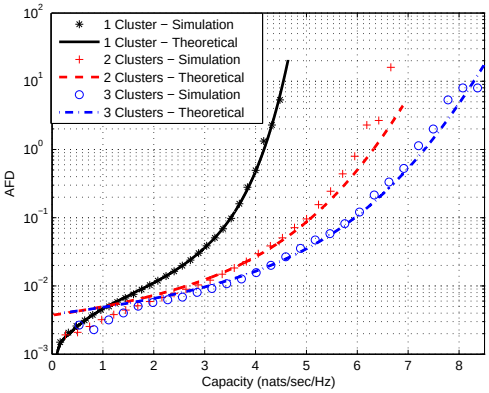
(a) LCR - 8x8 MIMO



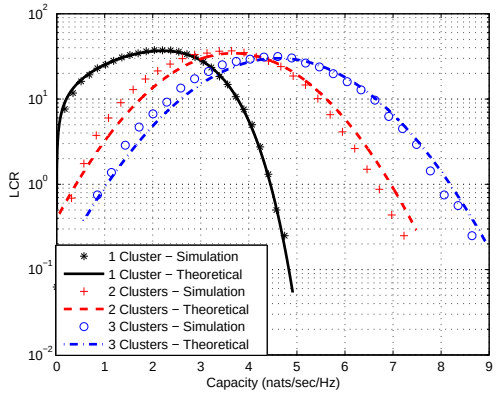
(b) AFD - 8x8 MIMO



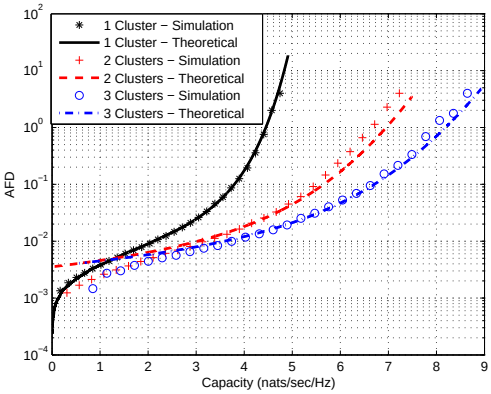
(c) LCR - 12x12 MIMO



(d) AFD - 12x12 MIMO



(e) LCR - 16x16 MIMO



(f) AFD - 16x16 MIMO

Figure 4.8: Level-Crossing Rate and Average Fade Duration of Capacity for Various Cluster Sizes

Tractable approximations to capacity in the case of multiple clusters is also provided and verified through numerical simulations. All quantities as derived play a vital role in design of adaptive transmission systems.

The application of findings in this chapter can be used in a number of ways. For example, in the next chapter, the SNR distribution of individual channels is used in design of a multi-queue scheduling algorithm. Furthermore, using the Wang-Moyaeri model, one can measure other useful system metrics such as how fast the channels evolve in time, or quantifying the duration of time the channel can be assumed static when applying block fading model assumptions. Details of future work on this chapter is described in Chapter 8.

Chapter 5
Time-Varying Energy Efficient Bit and Packet-Based Scheduling in Sparse Channels with CDI

In Chapter 4 we studied a particular class of MIMO channels where the channel is composed of a small number of finite clusters. In this chapter, we are interested in the distribution of the subchannel SNR and exploiting this knowledge in the scheduler design for the $\{K \times M\}$ system as described in Chapter 3. The benefits of optimization to meet QoS constraints was previously shown in Chapter 3 for the case of static. The purpose of this chapter is to propose a method to extend the work described in Chapter 3 to the case of time-varying channels as well as to incorporate the use of coded modulation schemes. We also propose several complexity reduction methods in the formulation of the optimization problem.

The remainder of this chapter is divided into five sections. Section 5.1 briefly outlines the medium access control (MAC) and physical (PHY) layer models of the downlink system. The MAC model was discussed in detail in Section 3.1 of Chapter 3. In this chapter, we relax the previous assumption that the subchannels are static in time. It is assumed that knowledge of only the channel distribution information (CDI) is exploited in the scheduler policy design since the policy is computed offline[1]. Next, in Section 5.2 we detail the overall time-varying scheduler design. Next, in Section 5.3, the PHY layer allocation mechanism is presented and formulated using general optimization framework. Results comparing the average arrival rate and

1. While only CDI is used in the scheduling policy determination, full channel state information (CSI) is assumed at the transmitter while transmitting using this policy online. We note that full CSI is required for MIMO singular value decomposition (SVD).

delay tradeoff on average allocated power is then shown in Section 5.4 along with discussion on tradeoffs with the number of channel partitions in the PHY. Finally, a brief summary is drawn in Section 5.5. Part of this work appears in Dechene et al. [8].

5.1 System Model and Overview

The multi-queue, multi-channel downlink system is composed of a base station transmitting traffic to a single subscriber station as shown in Figure 3.1 in Chapter 3. The system operates is divided into a MAC component and PHY component. The MAC component schedules packets subject to QoS constraints, queue occupancy information and transmission cost information from the PHY layer. The PHY layer is responsible for performing channel, rate and power allocation subject to channel state information and instantaneous packet service rates to minimize transmission cost while satisfying channel loss constraint. This MAC model was described in detail in Chapter 3 and summarized below while the modified PHY model to account for the time-varying channel is presented here.

There are K traffic classes using independent FIFO buffers with parameter set $\{\mathcal{D}_i, L_i, \overline{\lambda}_i, B_i, \delta_i\}$ describing the maximum tolerable average delay, packet size, average arrival rate, buffer size and maximum tolerable packet loss rate respectively. The tolerable loss rate give a packet arrives to the queue can further be expressed by its components: $\delta_i = 1-(1-P_{d,i})(1-P_{l,i})$; where $P_{d,i}$ and $P_{l,i}$ denote the probability of packets dropped entering the queue and packets dropped due to channel errors respectively. Similar to Chapter 3, it is assumed that these losses are both known and constrained independently. During each frame, a number of packets are taken from each queue and transmitted to the subscriber station. The scheduling algorithm employed is QoS-aware and formulates scheduling decisions based on the expected transmission cost and information about its instantaneous buffer levels and individual traffic class QoS requirements.

The scheduling time horizon is divided into a number of frames as shown in Figure 5.1. Each frame has a duration of T_f seconds and has a fixed number of symbols

of duration T_s seconds in each subchannel. There are a number of transmission rates that can be chosen which is referred to as the modulation and coding scheme (MCS). For each possible MCS, each subchannel can carry a given number of bits. From frame to frame, the evolution of each buffer can be described by the number of departures $c_i(n)$, the number of arrivals $A_i(n)$ and the previous buffer occupancy level $u_i(n)$. An example frame allocation is shown in Figure 5.1. Here, we see four streams being transmitted over four subchannels.

The base station and receiver are both equipped with multiple antennas. There are M_T antennas at the base station and M_R antennas at the receiving station. The system employs singular value decomposition (SVD) eigenbeamforming. In general a MIMO SVD system allows up to $M = \min\{M_T, M_R\}$ parallel subchannels for transmission. Measurement campaigns however have suggested [55, 56] that the number of non-zero eigenvalues in general is less than the minimum number of antennas (i.e., $L \leq \min\{M_T, M_R\}$). Such channels are known as sparse, and in fact can be modeled using a well-described geometric approach [65] to which the underlying time-varying parameters were well-studied earlier in this thesis and are described in Chapter 4. It is assumed that in this chapter, L is known.

From Chapter 4, we know the ℓ^{th} unordered subchannel eigenvalue has a known

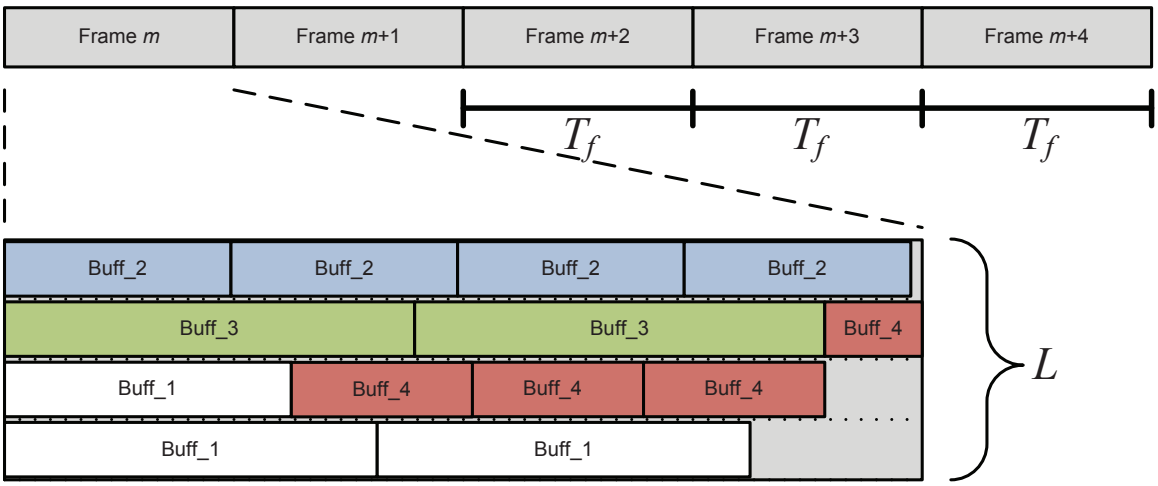


Figure 5.1: Frame Timing Layout for Packet Allocation

probability density function (PDF) of[2]

$$f_{\Lambda_\ell}(\lambda_\ell) = \frac{1}{M_R M_T P_\ell} \exp\left(-\frac{\lambda_\ell}{M_R M_T P_\ell}\right) \tag{5.1}$$

where P_ℓ is the relative power of that contributing cluster (described in [65]) such that $\sum_{\ell=1}^{L} P_\ell = 1$. The equivalent SNR of each subchannel of the SVD system for a reference SNR γ_0 is then [67]

$$\gamma_\ell = \frac{\gamma_0}{M_T}\lambda_\ell \tag{5.2}$$

Finally, the density function of the subchannel at a reference SNR γ_0 is simply

$$f_{\Gamma_\ell}(\gamma_\ell) = \frac{1}{\gamma_0 M_R P_\ell} \exp\left(-\frac{\gamma_\ell}{\gamma_0 M_R P_\ell}\right) \tag{5.3}$$

While each time-varying subchannel SNR can be described by the distribution in (5.3), recent cross-layer design techniques [20, 21] partition the subchannels into a finite number of states [19] in scheduler design which helps to reduce decision complexity by reducing the number of states needed to describe the system.

In the case of multiple parallel subchannels, the overall state of the channel can be described by jointly considering the state of all parallel subchannels. Each subchannel is divided into $\mathbb{J}_\ell + 1$ states; $\mathbb{J}_\ell$ active states as well as 1 null (or outage state). It is assumed that no transmission occurs in a subchannel in an outage state. The impact and existence of an outage state is discussed later in this chapter. Denoting $\mathcal{J}_\ell$ as a set of $\mathbb{J}_\ell$ finite active states of the ℓ^{th} unordered subchannel SNR, each j_ℓ^{th} subchannel state is bounded by $[\varphi_{\ell,j_\ell}, \varphi_{\ell,j_\ell+1})$ where $\varphi_{\ell,1} = \varepsilon$, and $\varphi_{\ell,\mathbb{J}_\ell+1} = \infty$ for each subchannel ℓ. The 0^{th} state bounded by $[0, \varepsilon)$ denotes the null state where ε is a small positive number chosen so the probability of outage is near 0 ($\Pr[j_\ell = 0] \approx 0$). Further, we can express $\mathcal{J}$ (or the joint subchannel state) as

$$\mathcal{J} = \mathcal{J}_1 \times \cdots \times \mathcal{J}_L \tag{5.4}$$

2. While we focus on the illustrative case of sparse MIMO channels, extensions are trivial for any multichannel model with known SNR distribution. As a result the presented framework can be applied in a general multiple channel scenario.

There are a number of partitioning methods discussed in the literature [19, 20]. Further research has shown [21] that properly designed channel partitioning thresholds can offer another degree of design freedom in adaptive transmission design, however we do not focus on that here[3]. Here, we employ an equal probability method such that the bounds are chosen to satisfy

$$\int_{\varphi_{\ell,j_\ell}}^{\varphi_{\ell,j_\ell+1}} f_{\Gamma_\ell}(r)dr = \frac{1}{\mathbb{J}_\ell}, \qquad \begin{matrix} j_\ell = 1, 2, \ldots, \mathbb{J}_\ell \text{ and} \\ \ell = 1, 2 \ldots, L \end{matrix} \tag{5.5}$$

where for a given number of partitions $\mathbb{J}_\ell$ and for a subchannel SNR distribution in (5.3), the above can be easily found as

$$\varphi_{\ell,j_\ell+1} = -\gamma_0 M_R P_\ell \ln\left(1 - \frac{j_\ell}{\mathbb{J}_\ell}\right), j_\ell = 1, \ldots, \mathbb{J}_\ell - 1 \tag{5.6}$$

Consider the following example where both the base station and subscriber station are equipped with 8 antennas, a reference SNR of $\gamma_0 = 10$dB and the channel is dominated by 4 independent scatters with relative powers $P_\ell = \{0.4, 0.3, 0.2, 0.1\}$. Partitioning each eigenvalue into 5 active states and using the equal probability method described above, the probability of being in a certain active state to equal $1/5 = 0.2$. As such each boundary can be given as:

$$\varphi_{\ell,j_\ell+1} = -\gamma_0 M_R P_\ell \ln\left(1 - 0.2 j_\ell\right), j_\ell = 1, \ldots, \mathbb{J}_\ell - 1 \tag{5.7}$$

and for the example above, all boundaries are given in Table 5.1.

The power level required to maintain a given loss rate on the channel, $P_{l,i}$, will depend on both the subchannel states described above, the subchannel of interest, the packet size and chosen MCS mode in that subchannel of interest. By using packet level allocation, packet error rates for coded transmissions can be easily accounted for using known analytical expressions. Here, we use the block outage probability

3. We note that while the number of partitions can impact the scheduler computational complexity, the method of choosing partition thresholds of the channel does not.

Table 5.1: Channel Partition Boundaries of FSMC Model

Boundaries	$\ell = 1$	$\ell = 2$	$\ell = 3$	$\ell = 4$
$\varphi_{\ell,0}$	0	0	0	0
$\varphi_{\ell,1}$	ε	ε	ε	ε
$\varphi_{\ell,2}$	7.141	5.355	3.570	1.785
$\varphi_{\ell,3}$	16.346	12.260	8.173	4.087
$\varphi_{\ell,4}$	29.321	21.991	14.661	7.330
$\varphi_{\ell,5}$	51.502	38.627	25.751	12.876
$\varphi_{\ell,6}$	∞	∞	∞	∞

(BLOP) derived in [10] to model the instantaneous packet error rate of the subchannel which is given as

$$PER(\gamma, k_\ell, L_i) \approx Q\left(\frac{\log(1+\gamma) - \log(2)k_\ell}{\sqrt{\frac{2k_\ell}{L_i}\frac{\gamma}{1+\gamma}}}\right) \tag{5.8}$$

where γ is a given SNR level, k_ℓ is the spectral efficiency (number of bits per symbol) and $Q(\cdot)$ is the Q-function. The factor $\log(2)$ is reflection of the fact that we measure spectral efficiency in bits per symbol instead of nats/symbol. The above was shown to closely approximate the block error rate (BLER) for medium block lengths such as those used later.

Therefore the loss rate on the subchannel is given as the average PER over active state j_ℓ of subchannel ℓ with applied power $\mathcal{P}$ as

$$P_{l,i} = \frac{1}{|\mathcal{J}_\ell|} \int_{j_\ell}^{j_\ell+1} PER(\gamma_\ell \mathcal{P}, k_\ell, L_i) f_{\Gamma_\ell}(\gamma_\ell) d\gamma_\ell \tag{5.9}$$

The value of $\mathcal{P}$ satisfying the above for a given $P_{l,i}$ in queue i in subchannel ℓ in state j_ℓ with a spectral efficiency of k_ℓ is denoted $P(i, \ell, j_\ell, k_\ell)$. For any outage state $P(i, \ell, 0, k_\ell), \forall i, \ell, k_\ell$, the solution $\mathcal{P}$ evaluates to infinity. For any subchannel in an outage state (*i.e.,* $j_\ell = 0$) there is no transmission in that subchannel therefore the only allowable spectral efficiency is 0 (*i.e.,* $k_\ell = 0$).

Table 5.2: Queue Configuration Parameters

	D_i	L_i	$\bar{\lambda}_i$	B_i	δ_i
Queue 1	3	200	2	25	0.05
Queue 2	5	400	1	30	0.01
Queue 3	5	250	2	25	0.10

Now suppose the system has 3 queues over the physical layer example previously described with parameter set $\{\mathcal{D}_i, L_i, \overline{\lambda}_i, B_i, \delta_i\}$ with quantities measured in frames, bits, packets per frame, packets and percentage respectively and values given in Table 5.2. Further, suppose during any frame n, the scheduler can take between 0 and 4 packets from each queue. Suppose valid spectral efficiencies are of $\{0, 1.5, 3, 4.5, 6\}$ bits per symbol. Furthermore, without loss of generality, T_f is normalized such that $T_f = 1$ and the symbol duration is $T_s = 0.005T_f$. In this case, the maximum data rate of the channel is then given as $6 \times 4 \times T_f/T_s = 4800$ bits/T_f which is less than our maximal required datarate of $(200 + 400 + 250) \times 4 = 3400$ bits/T_f . Finally suppose the maximal tolerable packet loss rate on the channel is half the total loss rate ($P_{l,i} = 0.5\delta_i$). We can then compute the above for each i, ℓ, j_ℓ, k_ℓ. Due to the monotonicity of (5.8) in $\mathcal{P}$ for fixed parameters, the above can be solved by numerical bisection. For example, for $i = 2$, $\ell = 3$, $j_\ell = 2$, $k_\ell = 4.5$ one obtains $P(2, 3, 2, 4.5) = 6.6494$. Similar values can be obtained for all other values of i, ℓ, j_ℓ, and k_ℓ.

5.2 Scheduler Design and Formulation

The scheduler has two-stages originally described in Chapter 3 where the scheduler allocates a set of MAC rates (queue service rates) based on the QoS parameters (*i.e.,* delay, throughput and buffer occupancy levels), and performs power, rate and channel allocation based on the required MAC rate in conjunction with channel CSI. In this chapter, we propose a modified algorithm for the power, rate and channel allocation stage, for use in conjunction with the MAC rate assignment stage we presented in

Chapter 3. The benefits of the proposed method in this chapter is the extension to time-varying channels, as well as the ability to utilize coded transmissions.

Due to physical transmission limitations, only a small number of packets relative to the queue size can be serviced from the queue during each frame. Let $\mathcal{C}_i$ be the set of possible MAC rates (*i.e.,* a set containing the possible quantity of packets that can be serviced from queue i during any frame). Subsequently, $\mathcal{C}$ can then describe all possible combinations of queue service rates across the set of queues or equivalently as

$$\mathcal{C} = \prod_{i=1}^{K} \mathcal{C}_i \tag{5.10}$$

Given a set $\mathcal{C}$ expressing the exhaustive MAC transmission rates for the MAC layer, each $c \in \mathcal{C} = \{c_1, c_2, \ldots, c_K\}$, target channel losses $P_{l,i}$, the set of channel states $\mathcal{J}$ and the set of valid MCS modes $\mathcal{M}$, the physical layer allocation scheme can be formulated as follows.

First, one can express the channel state and MAC rate assignment state space as $\mathcal{S} = \mathcal{C} \times \mathcal{J}$. For each $s \in \mathcal{S}$, the problem is formulated as follows.

To select a set of MCS modes $\mathbf{k}(s) = \{k_\ell(s), \forall \ell\}$ and a channel mapping scheme $\bar{X}_{\ell,i}(s), \forall \ell, i$ such that the total average power level selection is minimized. Let $\bar{X}_{\ell,i}(s)$ denote the number of packets from queue i to be transmitted in subchannel ℓ while in system state s with a transmission power of $P(i, \ell, \mathcal{J}_\ell(s), k_\ell(s))$, and where $\mathcal{J}_\ell(s)$ is the state of subchannel ℓ while in system state s. The power level $P(i, \ell, \mathcal{J}_\ell(s), k_\ell(s))$ is found from solving (5.9) for $\mathcal{P}$ for a given target channel loss rate $P_{l,i}$. Due to the monotonicity of (5.8) in γ, $\mathcal{P}$ can be found from (5.9) using efficient numerical techniques.

With the above definition, we define the power, channel and MCS allocation optimization (for $s \in \mathcal{S}$) subproblem to minimize the average applied transmission power as

$$\bar{P}(s) = \min_{\mathbf{k}(s), \bar{\mathbf{X}}(s)} f(\mathbf{k}(s), \bar{\mathbf{X}}(s)) \tag{5.11}$$

where

$$f(\mathbf{k}(s), \bar{\mathbf{X}}(s)) = \sum_{i=1}^{K} \sum_{\ell=1}^{L} \sum_{k_\ell \in \mathcal{M}}^{M} S_{k_\ell,\ell}(s) \cdot P(i, \ell, \mathcal{J}_\ell(s), k_\ell) \left\lceil \frac{\bar{X}_{\ell,i}(s) L_i}{k_\ell} \right\rceil \tag{5.12}$$

where $\mathbf{k}(s)$ and $\bar{\mathbf{X}}(s)$ are the vector and matrices containing $k_\ell(s), \forall \ell$ and $\bar{X}_{\ell,i}(s), \forall \ell, i$ respectively, $k_\ell(s)$ is the spectral efficiency of the chosen MCS mode in bits per symbol and $\lceil \cdot \rceil$ denotes rounding up to the nearest integer (ceiling function). The quantity $S_{k_\ell(s),\ell}(s)$ is an indicator function such that $S_{k_\ell(s),\ell}(s) = 1$ if a particular MCS mode k_ℓ is used for transmission in subchannel ℓ and 0 otherwise. Therefore $k_\ell(s) = \{k_\ell | S_{k_\ell,\ell}(s) = 1\}$. Further, we have the following additional constraints

$$\sum_{\ell=1}^{L} \bar{X}_{\ell,i}(s) = c_i(s), \quad \forall i \tag{5.13}$$

$$\sum_{i=1}^{K} \bar{X}_{\ell,i}(s) L_i \leq \frac{k_\ell(s) T_f}{T_s}, \quad \forall \ell \tag{5.14}$$

$$\sum_{k_\ell \in \mathcal{M}} S_{k_\ell,\ell}(s) = 1, \quad \forall \ell \tag{5.15}$$

$$\bar{X}_{\ell,i}(s) \in \mathbb{I}, \quad \forall i, \ell \tag{5.16}$$

$$\bar{X}_{\ell,i}(s) \geq 0, \quad \forall i, \ell \tag{5.17}$$

$$S_{k_\ell,\ell}(s) \in \{0, 1\}, \quad \forall \ell \tag{5.18}$$

The constraint in (5.13) ensures that the number of allocated packets for each stream across all subchannels satisfies the MAC requested rate $c_i(s)$, while (5.14) ensures that the selected MCS mode in each subchannel satisfies the amount of data transmitted over that subchannel. The constraint in (5.15) enforces that only a single MCS mode can be applied per subchannel during a given time frame. Finally, constraints (5.16)-(5.18) enforce integer/binary restrictions and non-negativity constraints on $\bar{X}_{\ell,i}(s)$ and $S_{k_\ell,\ell}(s)$.

Solutions to the above problem requires enumeration of all possible combinations of $S_{k_\ell,\ell}(s)$. This scales as $\mathcal{O}(|\mathcal{M}|^L)$ and results in a large number of possible

MCS mode combinations for the optimization routine above. To improve computation efficiency, we take the following two steps in the pre-solution stage. First, the constraints given from (5.13)-(5.18) suggest that only a subset of eligible candidate sets $\{S_{k_\ell,\ell}(s), \forall \ell\}$ that satisfy the constraints exists. The second is by using subchannel ordering (assigning the highest rate to that of the best quality channel and so on).

5.2.1 MCS Selection Space Reduction

The MCS space reduction is as follows. For each $s \in \mathcal{S}$, let $\mathcal{K}$ be the set of all MCS mode combinations by enumerating each possible combination of $S_{k_\ell,\ell}(s)$. Further, let $\mathcal{K}_{inf}(s)$, $\mathcal{K}_e(s)$, and $\mathcal{K}_o(s)$ be non-overlapping subsets of $\mathcal{K}$ such that

$$\mathcal{K} = \mathcal{K}_{inf}(s) \bigcup \mathcal{K}_e(s) \bigcup \mathcal{K}_o(s) \tag{5.19}$$

Here $\mathcal{K}_{inf}(s)$ is the set of MCS modes assignments that are infeasible due to insufficient sum rate (fails to satisfy (5.14))

$$\mathcal{K}_{inf}(s) = \left\{ \mathbf{k} \in \mathcal{K} \left| \sum_{\ell=1}^{L} k_\ell \leq \llcorner \epsilon_{inf}(s) \lrcorner, \right. \right\} \tag{5.20}$$

where

$$\epsilon_{inf}(s) = \frac{T_s}{T_f} \sum_{i=1}^{K} c_i(s) L_i \tag{5.21}$$

with $\llcorner \cdot \lrcorner$ denoting rounding down to the next smallest possible MCS mode combination for a given set of allowable spectral efficiencies.

The subset $\mathcal{K}_o(s)$ defines what we refer to as the *overfeasible* set. This is the subset of MCS mode combinations that likely do not yield the most energy efficient allocation. In general, this set is not unique. By maximizing the size of the *overfeasible* set, one can minimize the system complexity (i.e., the number of optimization subproblems). We strongly suggest that the largest obtainable *overfeasible* set without

eliminating the most energy efficient allocations is given as

$$\mathcal{K}_o(s) = \left\{ \mathbf{k} \in \mathcal{K} \left| \sum_{\ell=1}^{L} k_\ell > \ulcorner \epsilon_o(s) \urcorner, \mathbf{k} = \{k_1, \ldots, k_L\} \right. \right\} \tag{5.22}$$

where

$$\epsilon_o(s) = \frac{T_s}{T_f} \left(L \max\{L_1, \ldots, L_K\} + \sum_{i=1}^{K} c_i(s) L_i \right) \tag{5.23}$$

and $\ulcorner \cdot \urcorner$ denotes rounding up to the next valid MCS mode combination for a given set of allowable spectral efficiencies. The argument for the above is as follows. First, from (5.20) and by the monotonically increasing nature of (5.8) in k_ℓ, increasing the spectral efficiency of any single channel beyond what is required for channel allocation is inefficient (requires an increase in power to maintain a target PER). Combining this with the granularity of the problem (packet-level assignment granularity), it is possible that the most energy efficient MCS mode selection scheme must be able to assign up to 1 more of the largest granular quantity (largest packet) into any subchannel ℓ. Finally since $\mathcal{K}_o(s)$, $\mathcal{K}_e(s)$, and $\mathcal{K}_{inf}(s)$ are non-overlapping sets satisfying (5.19), one can find $\mathcal{K}_e(s)$ as[4]

$$\mathcal{K}_e(s) = (\mathcal{K} \setminus \mathcal{K}_{inf}(s)) \setminus \mathcal{K}_o(s) \tag{5.24}$$

where $\mathcal{K}_e(s)$ is the set possible MCS mode combinations used below.

4. Here we attempt to clarify a couple points to the reader. Firstly, there may exist (due to channel mapping granularity restrictions) modes that propose infeasible solutions to the optimization subproblem, and secondly, not all modes in $\mathcal{K}_e(s)$ yield the most energy efficient MCS mode selection. We emphasize the purpose of this set is to determine a small subset of all possible MCS modes to perform the above optimization in order to reduce the search space in advance.

5.2.2 MCS Assignment and Subchannel Ordering

The subchannel ordering operation is as follows. First, the mean SNR value of subchannel ℓ in state $\mathcal{J}_\ell(s)$ is given by

$$\mu_\ell(s) = \int_{\mathcal{J}_\ell(s)} \gamma_\ell f_{\Gamma_\ell}(\gamma_\ell) d\gamma_\ell \tag{5.25}$$

Each $\mathbf{k} \in \mathcal{K}_e(s)$ vector contains the MCS modes for all L subchannels. Let $\bar{\mathbf{k}}$ represent the ordered vector $\mathbf{k}$ such that $\bar{k}_1 \geq \cdots \geq \bar{k}_x \geq \cdots \geq \bar{k}_L$. Let $x = \mathfrak{F}(\ell)$ (with corresponding inverse function $\ell = \mathfrak{F}^{-1}(x)$) return the subchannel level rank x of subchannel ℓ. A subchannel level rank of x means that subchannel ℓ has the x^{th} highest mean as given by (5.25) (*i.e.*, $\mu_{\mathfrak{F}^{-1}(1)}(s) \geq \cdots \geq \mu_{\mathfrak{F}^{-1}(\ell)}(s) \geq \cdots \mu_{\mathfrak{F}^{-1}(L)}(s)$). Each MCS mode $\bar{k}_x$ is mapped to each ordered subchannel $\mu_{\mathfrak{F}^{-1}(x)}(s)$.

5.2.3 MCS Mode Selection Example

In order to better understand derivation of the potentially energy efficient MCS mode subset, consider the following. Employing the example described earlier in this chapter and supposing that the system is solving for $s \in \mathcal{S}$ and that the respective MAC rates require transmission of 2 packets from queue 1, 4 from queue 2 and 1 from queue 3.

5.2.3.1 Infeasible Set

First to determine the infeasible set from (5.20) we have

$$\epsilon_{inf}(s) = \frac{0.005}{1}(2 \cdot 200 + 4 \cdot 400 + 1 \cdot 250) = 11.25 \tag{5.26}$$

where we can see from (5.20) that any set of MCS modes $\mathbf{k}$ such that $\sum_{\ell=1}^{L} k_\ell(s) \leq \llcorner 11.25 \lrcorner$. Supposing valid spectral efficiencies are $\{0, 1, 2, 4, 6\}$ corresponding to no transmission, BPSK, QPSK, 16QAM and 64QAM respectively (uncoded), it can easily be shown that $\llcorner 11.25 \lrcorner = 11$.

Table 5.3: Potentially Energy Efficient MCS Mode Sets

C_1	C_2	C_3	C_4	C_1	C_2	C_3	C_4
6	6	0	0	6	4	2	0
4	4	4	0	6	4	1	1
6	2	2	2	4	4	2	2
6	6	1	0	6	4	2	1
4	4	4	1	6	6	2	0
6	4	4	0	6	6	1	1
6	4	2	2	4	4	4	2
6	6	2	1	6	4	4	1
6	6	4	0	6	6	2	2
6	4	4	2	4	4	4	4
6	6	4	1	6	6	6	0
6	6	4	2	6	4	4	4

5.2.3.2 *Overfeasible* Set

Similarly from (5.22) one can find MCS modes combinations that exist in the suggested *overfeasible* set where

$$\epsilon_o(s) = \frac{0.005}{1}\left(4 \cdot 400 + 2 \cdot 200 + 4 \cdot 400 + 1 \cdot 250\right) = 19.25 \tag{5.27}$$

where again it is easy to show that $\lceil 19.25 \rceil = 20$.

5.2.3.3 Energy Efficient Set

Based on the above example and allowable spectral efficiencies, the potentially energy efficient set is given as the set of MCS modes yielding the following sum of spectral efficiencies:

$$\sum_{\ell=1}^{L} k_\ell(s) = \{12, 13, 14, 15, 16, 17, 18\} \tag{5.28}$$

From this, combined with MCS subchannel ordering, we can obtain a total of 24 potential ordered assignment modes detailed in Table 5.3 where C_ℓ denotes the ordered subchannel index such that $\mu_{C_1}(s) \geq \mu_{C_2}(s) \geq \mu_{C_3}(s) \geq \mu_{C_4}(s)$.

5.2.3.4 Complexity Reduction

In the above example, one can see the complexity reduction achieved using the proposed method. An exhaustive set of $\mathcal{S}_{k_\ell,\ell}$ for the above example suggests there are $6^4 = 625$ possible ways an MCS mode can be selected for each state $s \in \mathcal{S}$, however the resulting potentially energy efficient subset of MCS modes yields only 24 ordered sets. A resulting reduction in complexity by factor of 26 times through simple precomputation.

5.2.4 Revised Optimization Formulation

One can revise the transmission cost function from (5.12) to be a function of the allocation matrix $\mathbf{X}(s)$ as

$$f^{\bar{\mathbf{k}}(s)}(\bar{\mathbf{X}}(s)) = \sum_{i=1}^{K}\sum_{x=1}^{\bar{L}} P(i, \mathfrak{F}^{-1}(x), \mathcal{J}_{\mathfrak{F}^{-1}(x)}(s), \bar{k}_x(s)) \cdot \left\lceil \frac{\bar{X}_{\mathfrak{F}^{-1}(x),i}(s) L_i}{\bar{k}_x(s)} \right\rceil \tag{5.29}$$

where $\bar{L}$ is the number of subchannels where $k_\ell > 0$ (in use during a given state) and where the solution to the optimization problem follows as

$$\bar{P}(c,j) = \bar{P}(s) = \min_{\bar{\mathbf{k}}(s)\in\mathcal{K}_e(s)} \left(\min_{\bar{\mathbf{X}}(s)} f^{\bar{\mathbf{k}}(s)}(\bar{\mathbf{X}}(s)) \right) \tag{5.30}$$

For each $s \in \mathcal{S}$ (or $j \in \mathcal{J}$, $c \in \mathcal{C}$), the above can be solved in two stages since in our model the channel evolves independently of the action space, which results in significantly lower complexity than joint optimization. This is true in this formulation as the queue action determination is based on the long-term average allocation, and not the instant of time. Finally, $\bar{P}(\mathbf{c})$, used in (3.24) in Chapter 3, can be found by averaging over all possible subchannel states or as

$$\bar{P}(\mathbf{c}) = \frac{1}{|\mathcal{J}|} \sum_{j\in\mathcal{J}} \bar{P}(c,j) \tag{5.31}$$

where $|\cdot|$ denotes the size of a set. As before, the above framework can solve for all quantities offline and in advance, where the resulting resource allocation quantities can be stored in a look up table (LUT) at the base station. Full details of the LUT were described in Chapter 3.

5.3 Formulation of Optimization Problem

The channel, rate and power allocation described in the previous section can be formulated using the well-known branch and bound technique. The branch and bound algorithm can be used to solve a LP problem where one or more components of the solution vector are integers. A general branch and bound problem is formulated to solve $f(\mathbf{x}) = \arg\min_{\mathbf{x}} \mathbf{c}^T\mathbf{x}$ subject to $\mathbf{A}\mathbf{x} \leq \mathbf{b}$, $\mathbf{A}_{eq}\mathbf{x} = \mathbf{b}_{eq}$ and $\mathbf{x} \geq 0$ where $\mathbf{A}$ and $\mathbf{A}_{eq}$ are matrices, $\mathbf{b}$, $\mathbf{b}_{eq}$ and $\mathbf{c}$ are vectors and some or all entries in $\mathbf{x}$ are constrained to integers.

The vector $\mathbf{x}$ is a $\bar{L}K \times 1$ vector with integer elements $\bar{X}_{\mathfrak{F}^{-1}(x),i}, x = 1, 2, \ldots, \bar{L}, i = 1, 2, \ldots, K$ given as

$$\mathbf{x} = [\bar{X}_{\mathfrak{F}^{-1}(1),1}(s),\ \ldots,\ \bar{X}_{\mathfrak{F}^{-1}(1),K}(s),\ \ldots,\ \bar{X}_{\mathfrak{F}^{-1}(\bar{L}),K}(s)]^T \tag{5.32}$$

The objective function from (5.29) can be described as a coefficient vector $\mathbf{c}$ with entries

$$\mathbf{c} = [\zeta_{\mathfrak{F}^{-1}(1),1}(s), \ldots, \zeta_{\mathfrak{F}^{-1}(1),K}(s), \zeta_{\mathfrak{F}^{-1}(2),1}(s), \ldots, \ldots, \zeta_{\mathfrak{F}^{-1}(\bar{L}),K}(s)] \tag{5.33}$$

where $s \in \mathcal{S}$ and $\zeta_{\ell,i}(s)$ is

$$\zeta_{\ell,i}(s) = P(i, \ell, \mathcal{J}_\ell(s), k'_{\mathfrak{F}(\ell)}) \left\lceil \frac{L_i}{k'_{\mathfrak{F}(\ell)}} \right\rceil \tag{5.34}$$

Table 5.4: Simulation Parameters

Parameter	**Value**
Number of Antennas (M_T, M_R)	8
Valid Spectral Efficiencies ($\mathcal{M}$)	$\{0,\ 1.5,\ 3,\ 4.5,\ 6\}$
Length of Packet (L_i bits)	200
Arrival Rate ($\bar{\lambda}_i$ packets/frame)	2
Buffer Size (B_i packets)	25
Average Packet Delay ($\mathcal{D}_i$ frames)	4
Total Loss Rate (δ, % of Packets)	10%
Target Channel Loss Rate ($P_{l,i}$)	$\delta/2$
Symbols per Frame per Channel (T_f/T_s)	200
MAC Rates ($\mathcal{C}$, packets/frame)	$\{0, 1, 2, 3\}$
Number of independent scatters (L)	4
Scatter relative power (P_l)	$\{0.4\ \ 0.3\ \ 0.2\ \ 0.1\}$
Number of active subchannel partitions ($\mathbb{J}_\ell$)	5
Reference SNR (γ_0)	10dB
ε	10^{-6}

5.3.1 Equality Constraints

The K equality constraints from (5.13) are given in the $K \times \bar{L}K$ matrix $\mathbf{A}_{eq}$ with entries

$$A_{eq:i,z} = \begin{cases} 1, z \in \mathcal{I}_i \\ 0, otherwise \end{cases} \tag{5.35}$$

where $\mathcal{I}_i$ is the set containing location indices of $\bar{X}_{\ell,i}, \forall \ell$ in $\mathbf{x}$. The coefficient vector $\mathbf{b}_{eq}$ is subsequently given as

$$\mathbf{b}_{eq} = [c_1(s)L_1,\ c_2(s)L_2, \ldots,\ c_K(s)L_K]^T \tag{5.36}$$

where $c_i(s)$ is the number of packets taken from queue i when the system is in state s.

5.3.2 Inequality Constraints

The $\bar{L}$ equality constraints from (5.14) are defined in the $\bar{L} \times \bar{L}K$ matrix $\mathbf{A}$ with entries

$$A_{x,z} = \begin{cases} 1, z \in \mathcal{I}'_x \\ 0, otherwise \end{cases} \tag{5.37}$$

where $\mathcal{I}'_x$ is the set containing location indices of $\bar{X}_{\mathfrak{F}^{-1}(x),i}, \forall i$ in $\mathbf{x}$. The vector $\mathbf{b}$ is given as

$$\mathbf{b} = \frac{T_f}{T_s}[k'_1,\ k'_2, \ldots,\ k'_{\bar{L}}]^T \tag{5.38}$$

Combining the above constraints, the problem can be computed through application of the branch and bound technique, which is described in Appendix B.

5.4 Numerical Evaluation

Selected simulation results are presented using parameters as given in Table 5.4. As in Chapter 3 the proposed framework is able to meet delay and throughput requirements. In this section we solely present results demonstrating the total average power level selection as well as the effect of subchannel partitioning in terms of complexity compared to average allocated power. These results are for a single user to isolate the direct tradeoffs of interest, however in all scenarios the proposed method was able to meet various QoS requirements of multiple traffic streams.

5.4.1 Average Power Usage

In Figure 5.2 we show the average power performance as a function of the delay constraint and arrival rate in using our newly extended PHY with the MAC described in Chapter 3. As expected, the power performance is largely dominated by the average arrival rate. Further, and as expected, the average power is related to the delay in that an increase in the delay tolerance reduces the total average power (albeit at a lesser impact than the average arrival rate). This is a result of the system being able to exploit channels higher SNR states, for higher rate transmission. Further, we

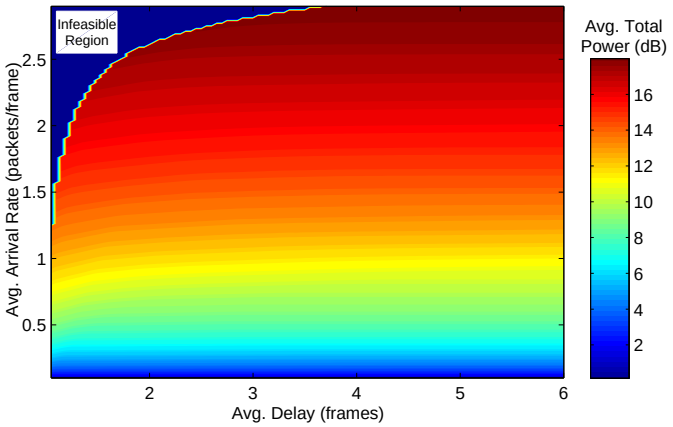


Figure 5.2: Total Average Power vs. Delay and Arrival Rate

see that a region of arrival rate/delay constraints is infeasible as a result of the set of MAC service rates employed. This is also consistent with well-known queueing theory stability results. A major observation above, is that while the impact on average power performance is impacted by the delay constraint, a larger impact is a result of the arrival rate.

5.4.2 Partitioning Performance

The total number of subchannel partitions has a large impact on system performance. In Figure 5.3 we show the impact on the number of partitions contrasted with the average power consumption. The computation time is measured with respect to the base case of 2 partitions per subchannel. On the one hand, we show that increasing the number of partitions increases the power efficiency of the system by allowing a greater degree of adaptability, however, the resulting complexity increases (resulting in increased computation time). Further, the memory requirement for the size of the lookup table greatly depends on the number of partitions. It is given as

$$SIZE_{LUT} = 64(2K+1)L\prod_{i=1}^{K}|\mathcal{C}_i|\prod_{\ell=1}^{L}(\mathbb{J}_\ell) \quad \text{bits} \tag{5.39}$$

where all relevant parameters are assumed to be stored as 64 bit doubles.

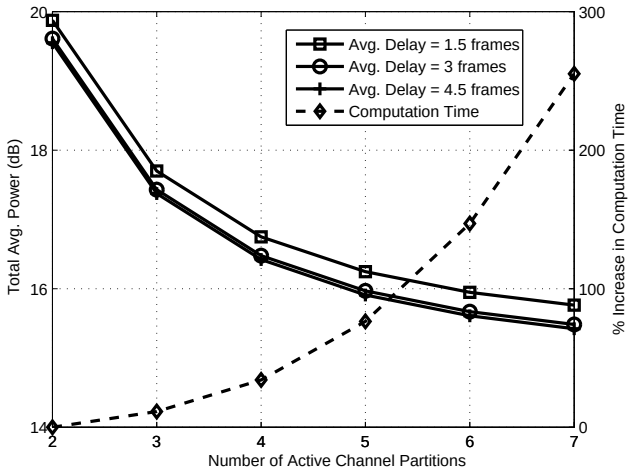


Figure 5.3: Impact Effect of Channel Partitioning on Average Applied Power

5.5 Chapter Summary

In this chapter, we extended the results of Chapter 3 to a packet based transmission scheme in addition to applying this framework to time-varying channels. This proposed framework proposes a method of incorporating of arbitrary MCS schemes in the optimization framework. In addition, we also study the effect of subchannel partitioning on resource allocation power and complexity.

There still remain a number of open problems and drawbacks to the system above. One large drawback is the small probability of an outage event. While by design the probability of this event is very small and has negligible impact on the overall queue performance in terms of delay and throughput, it is important to study a method to incorporate larger outage events in future work. Another drawback of the approach above is the disparity between channel information that can be exploited at the various stages of the scheduling policy. For example, for MIMO SVD, CSI is required at the transmitter, however only CDI in the form of average performance over all states of the channel is exploited in the queue service rate part of the policy. This presents an additional degree of flexibility that can be studied. Details of these, and other related open problems are discussed in future work given in Chapter 8.

Chapter 6
Energy Efficient HARQ-Aware SC-FDMA Uplink Resource Allocation

In previous chapters, we examined methods of scheduling multiple streams over multiple parallel channels, for downlink scenarios. Studying of the uplink scenario is also of importance, particularly in prolonging mobile battery lifetimes. Modern cellular technologies such as 3GPP-LTE utilize localized single carrier frequency division multiple access (SC-FDMA) in the uplink in order to share channel resources. SC-FDMA overcomes major issues for mobile devices associated with high peak to average power ratio (PAPR) when compared to traditional techniques such as OFDM. Unlike OFDM, localized SC-FDMA requires frequency resources to be allocated contiguously in frequency.

Focusing again on the multi-user uplink scenario from an abstracted view, a multi-user SC-FDMA channel is simply a multi-service, multi-channel system, where each user operates a single service and the individual frequency resources are available for allocation similar to parallel subchannels. In this case, the number of subchannels is large compared to previous examples using MIMO, and therefore exhibits large computational complexity. Moreover, with localized SC-FDMA, frequency resources must be allocated contiguously in frequency, unlike traditional methods such as OFDM.

In this chapter we look at the challenges presented by SC-FDMA with synchronous hybrid automatic repeat request (HARQ). Synchronous HARQ is a process used in LTE as it provides a mechanism to increase redundancy of transmissions while limiting overhead needed for asynchronous HARQ. Unfortunately, this introduces an additional complexity issue since the periodicity of this process needs to be considered by the scheduler since a user cannot be allocated new transmissions if there is an

ongoing retransmission in a subframe due to the HARQ process, which we refer to as *ARQ blocking*. To combat this problem, we propose a method of performing resource allocation that exploits this periodicity of the HARQ process in scheduler design. We propose use of a block time-frequency domain packet scheduler (BTFDPS). This proposed approach reduces the amount of scheduling decisions required for uplink traffic in time, in addition to simplifying incorporation of synchronous HARQ into the framework. At each block time-frequency frame, users are allocated contiguous frequency blocks, and the weighted average power allocation is minimized.

The remainder of this chapter is organized as follows. In Section 6.1 we overview the details of the employed uplink system model including the channel and scheduling models and in Section 6.2 we describe the scheduling ideology. In Section 6.3 we formulate the optimal energy efficient allocation problem with given HARQ constraints. Due to the complexity associated with the optimal allocation, we propose two sub-optimal methods in Section 6.4 in order to improve computational tractability. In Section 6.5 simulation results are provided to demonstrate the performance versus complexity of all three approaches while in Section 6.6, conclusions are drawn on this work. A portion of this chapter appeared in Dechene et al. [9].

6.1 System Model

The simplified MAC model of a multi-user SC-FDMA system used for uplink in 3GPP's LTE, [75] is shown in Figure 6.1. Here, each user has access to a single physical uplink shared channel (PUSCH) for transmission of their uplink data. Within a single cell there are K users (UEs) and a single base station (eNB). For the purpose of our work, it is assumed that intercell interference is negligible. The cell spectrum consists of N_{sub} narrowband subcarriers grouped into M resource blocks (RBs) of 12 subcarriers each. Without loss of generality, there is an integer number (M) of RBs. The system is assumed to be operating in frequency division duplexed (FDD) mode.

There are N_{sym} symbols per subcarrier in a given subframe where the exact number of subcarriers depends on the uplink configuration. The PUSCH is used for transmission of uplink data and shared between UEs (as shown in Figure 6.1). It is

assumed the PUSCH occupies $N_{sym} - N_{ctrl}$ symbols per subcarrier, per subframe where N_{ctrl} is the number of symbols used for uplink control data.

In 3GPP LTE, CQI values describe a range of targeted modulation and coding schemes (MCSs) and are given in Table 6.1 reproduces from [76]. The overall TB size is given as the effective spectral efficiency combined with the number of allocated RBs. Consequently, the overall number of data bits, the transport block (TB) size, that can be transmitted per subframe over a set $\mathcal{N}$ of RBs in frequency given the assumptions above is

$$\eta(b, \mathcal{N}) = \left\lfloor 12(N_{sym} - N_{ctrl})b \cdot |\mathcal{N}| \right\rfloor \tag{6.1}$$

where b is the spectral efficiency in bits per symbol, $\mathcal{N}$ is a set of RBs in frequency and $|\cdot|$ is the size of a set. For $N_{sym} = 14$ (regular cyclic prefix in LTE [76]) and assuming $N_{ctrl} = 3$, the above reduces to $\eta(b, \mathcal{N}) = \lfloor 132b \cdot |\mathcal{N}| \rfloor$. The minimum and maximum spectral efficiencies are found in Table 6.1 as $b_{min} = 0.1523$ and $b_{max} = 5.5547$.

Furthermore, the application of localized SC-FDMA constrains allocated RBs to any UE such that they are adjacent in frequency. Individual UEs are further limited to a single TB using a common MCS mode per subframe[1] and retransmissions employing HARQ must be transmitted over the same number of resource blocks using the same MCS after exactly 8 (ARQ_w) subframes have passed from original transmission (non-adaptive HARQ).

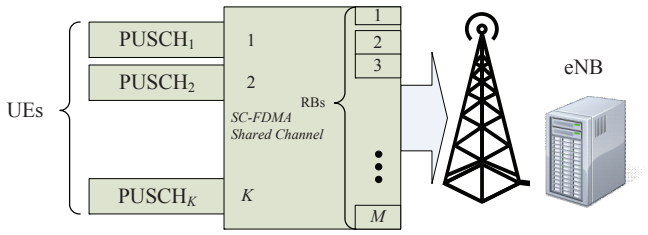


Figure 6.1: SC-FDMA Shared Channel Model

1. For single layer (non-MIMO) transmissions [76] which is assumed in this work.

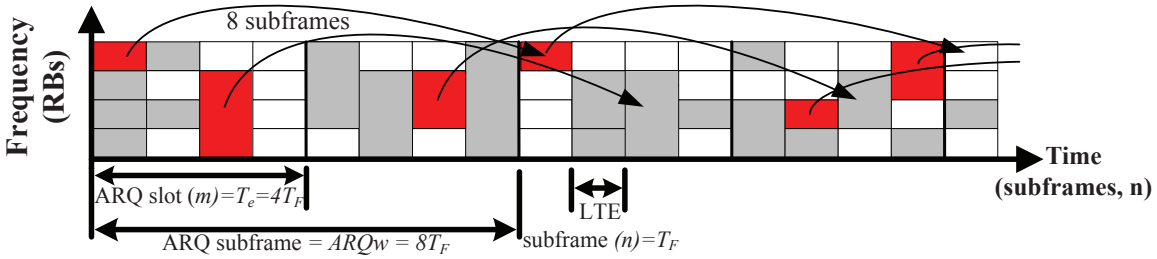


Figure 6.2: SC-FDMA Frame Layout with Synchronous HARQ

6.1.1 Scheduling Model

The minimum duration of time that can be allocated to a single UE is a scheduling block (SB). The n^{th} SB is in the time duration bounded by $[nT_f, (n+1)T_f)$ where T_f is the subframe duration (equal to 1ms) and is one RB in frequency.

Resource allocation decisions in LTE are made at each scheduling epoch. A scheduling epoch is the interval of time over which resource allocation decisions can

Table 6.1: List of CQI Indices (Modulation/Coding Schemes) [76]

Index	Modulation	Coding Rate	Spectral Efficiency - b in (6.1)
0	—	—	0 bits
1	QPSK	78/1024	0.1523
2	QPSK	120/1024	0.2344
3	QPSK	193/1024	0.3770
4	QPSK	308/1024	0.6016
5	QPSK	449/1024	0.8770
6	QPSK	602/1024	1.1758
7	16QAM	378/1024	1.4766
8	16QAM	490/1024	1.9141
9	16QAM	616/1024	2.4063
10	64QAM	466/1024	2.7305
11	64QAM	567/1024	3.3223
12	64QAM	666/1024	3.9023
13	64QAM	772/1024	4.5234
14	64QAM	873/1024	5.1152
15	64QAM	948/1024	5.5547

be made. While in general the scheduling epoch can be equal to the scheduling block (*i.e.*, $T_e = T_f$), we define the m^{th} scheduling epoch as the time duration in $[mT_e, (m+1)T_e)$ and each epoch consists of T_e/T_f subframes in time where m is the index of the scheduling epoch. The time T_e is chosen to be equal to

$$\frac{ARQ_w T_f}{TXOP} \tag{6.2}$$

where $TXOP$ denotes the maximum number of new TBs that can be allocated per ARQ_w subframes where

$$TXOP = \frac{ARQ_w}{Max_{tx}} \tag{6.3}$$

and Max_{tx} is the maximum number of times a packet can be (re)transmitted. The justification behind this selection T_e is discussed in later sections. Figure 6.2 shows an example of the uplink frame layout with relevant time durations shown. Successful transmission are shown in gray, while failures are red. As shown, retransmissions occur exactly ARQ_w subframes following the initial transmission using the same number of resource blocks.

We assume that each UE has a single radio data bearer established between itself and the core network. Each bearer has an access network bitrate of χ_i bit per unit time or an average newly allocated TB size of $\bar{T}_i$ bits per ARQ slot.

6.1.2 Channel Model

The channels between each UE and the eNB and from RB to RB are independent. A block fading model is assumed, where the channel is static for the duration of the decision epoch T_e and independent from epoch to epoch. Channel estimation is assumed to be error-free and available at the eNB and each channel follows the Rayleigh SNR distribution given as

$$p(\gamma) = \frac{1}{\gamma_0} \exp\left(-\frac{\gamma}{\gamma_0}\right) \tag{6.4}$$

where $\gamma_{i,k}(m)$ is the instantaneous SNR in resource block k seen from UE i during subframe m and it is average SNR γ_0 for a reference applied power level. where

$\gamma_{i,k}(m)$ is the instantaneous SNR will be used to denote the uplink channel of user i over RB k in epoch m.

The block error rate (BLER) is the average failure rate of TBs. This depends on $\gamma_{i,eff}$ (the effective signal to noise ratio, SNR, of a transmission), T_i (the TB size), and b_i (the effective MCS). To the best of our knowledge, there are no analytical solutions to compute the BLER of a general coded transmission. In practical implementations, estimates of BLER are obtained through receiver measurements. The measure of Information Outage probability (IOP) derived in [10] is used as a measure for BLER. The IOP was shown to closely model BLER for moderate block lengths. Any extensions to our proposed framework are trivial when estimates of the BLER can be more accurately obtained for a particular implementation using appropriate training and calibration techniques.

Following from [10], we can obtain the expression for BLER as

$$BLER(\gamma, \mathcal{N}_i, T_i) \approx Q\left(\frac{\log(1+\gamma) - \frac{\log(2)T_i}{132|\mathcal{N}_i|}}{\sqrt{\frac{2}{132|\mathcal{N}_i|}\frac{\gamma}{1+\gamma}}}\right) \tag{6.5}$$

where $\mathcal{N}_i$ is set of RBs allocated to UE i, $Q(\cdot)$ is the well-known Q-function and $\overline{\gamma}_{i,eff}$ is the effective SNR. The quantity $132|\mathcal{N}_i|$ and $\frac{T_i}{132|\mathcal{N}_i|}$ corresponds to the number of modulated symbols and effective spectral efficiency per allocation respectively described earlier. The factor $\log(2)$ is required and follows from [10] in that we measure spectral efficiency in bits per symbol rather than nats per symbol. Due to the monotonicity of the Q-function arguments, the above can be solved efficiently using bisection techniques for the required SNR. Alternatively, a more computationally efficient method is to obtain a least squares approximation to the above as a function of data rate, target BLER and the number of RBs allocated (similar to the approach in [20]). We found the following fitting function closes approximates the SNR as a function of data rate

$$\gamma_{i,eff}^{(r)} \approx a_x \exp(b_x T_i) - \gamma_{0,x} \tag{6.6}$$

where $x = |\mathcal{N}_i|$. The values of a_x, b_x, and $\gamma_{0,x}$ are given in Table 6.2 for $BLER_{tgt} = 10\%$ for up to 24RBs. Using the above, along with channel estimates, it is easy to

Table 6.2: Least-Squares Approximate Model Parameters for BLER=10% - Using Fit Function (6.6)

RBs (x)	a_x	$b_x \cdot 10^3$	$\gamma_{0,x}$
1	1.1748	5.2471	1.1019
2	1.1208	2.624	1.0723
3	1.0977	1.7495	1.0591
4	1.0841	1.3122	1.0512
5	1.0749	1.0498	1.0458
6	1.0682	0.8749	1.0418
7	1.063	0.7499	1.0387
8	1.0588	0.6562	1.0362
9	1.0553	0.5833	1.0342
10	1.0524	0.525	1.0324
11	1.0499	0.4772	1.0309
12	1.0478	0.4375	1.0296
13	1.0459	0.4038	1.0284
14	1.0441	0.375	1.0274
15	1.0426	0.35	1.0265
16	1.0412	0.3281	1.0256
17	1.04	0.3088	1.0249
18	1.0388	0.2917	1.0242
19	1.0378	0.2763	1.0235
20	1.0368	0.2625	1.0229
21	1.0359	0.25	1.0224
22	1.0351	0.2386	1.0219
23	1.0343	0.2283	1.0214
24	1.0336	0.2188	1.0209

obtain the required applied power to achieve a target block error rate. The accuracy of this approximation is verified in Section 6.5.

The measured effective SNR for a reference power level of a transmission is related to the SNR of the RBs comprising it. As in [77], the effective SNR of an SC-FDMA symbol cannot be approximated using EESM or MIESM (as in OFDM), but rather it can be approximated as the average SNR over the set of RBs allocated

or as

$$\gamma_{i,eff}^{(0)}(m,\mathcal{N}_i) = \frac{1}{|\mathcal{N}_i|}\sum_{k\in\mathcal{N}_i}\frac{\gamma_{i,k}(m)}{|\mathcal{N}_i|} \tag{6.7}$$

Here $\mathcal{N}_i$ is the set of RBs in the computation and the superscript (x) is used to denote the x^{th} retransmission number where 0 is the initial transmission. The additional $|\mathcal{N}_i|$ in the above equation describes that power is equally allocated across the all assigned resource blocks. We emphasize here that extension of the above is trivial where measurements/estimates of the effective SNR is obtained and known for a given implementation.

6.1.3 Retransmission Model

Synchronous HARQ is used for uplink retransmission to improve the probability that a TB can be decoded at the base station while limiting signalling overhead required for asynchronous HARQ. In the LTE implementation, a TB is first encoded with a 1/3 rate turbo code. From this, four possible codes blocks are obtained denoted as redundancy versions (RVs). Each redundancy version is a different set of coded bits from the TB. Each retransmission of a TB utilizes a different RV so that retransmissions can be combined with the initial transmission to improve overall reliability.

Consequently, this improved reliability can be considered as an increase in the measured effective SNR as done in [78] which is also used here. We employ this model with coefficients for SNR gains given in Table 6.3 but note that for a given implementation, these gain factors can be obtained through receiver calibration measurements. The effective SNR of retransmission z can be computed as (in dB)

$$\gamma_{i,eff,(dB)}^{(z)}(m,\mathcal{N}_i,k_i) = \gamma_{i,eff,(dB)}^{(0)}(m,\mathcal{N}_i) + \mu(z,k_i)\cdot R_{code} + \epsilon(z,k_i) \tag{6.8}$$

where R_{code} is the code rate $\times 1024$ which is given from [76] (*i.e.,* 78, 120, ...), $\mu(z,k_i)$ and $\epsilon(z,k_i)$ are given from [78] for each retransmission and modulation scheme, k_i is the modulation scheme and where $\gamma_{i,eff}^{(0)}(m,\mathcal{N}_i)$ is the effective SNR for UE i given in (6.7). These parameters are summarized in Table 6.3. The modulation scheme is

Table 6.3: SNR Gain for Retransmissions [78]

Modulation	rv_{idx}	$\mu \cdot 10^3$	ϵ
QPSK	1	0.8024	2.89
	2	1.628	4.57
	3	2.006	5.62
16-QAM	1	0.420	1.1
	2	8.435	0.74
	3	9.464	1.15
64-QAM	1	8.996	-1.23
	2	12.288	-0.71
	3	12.728	0.15

derived from the CQI mode in Table 6.1 yielding the next highest spectral efficiency mode than $\frac{T_i}{132|\mathcal{N}_i|}$.

6.1.4 Transmission Power Selection

For a given target block error rate (BLER) of $BLER_{tgt}$, the required power is simply

$$P_i(\mathcal{N}_i, T_i) = \frac{\overline{\gamma}_{i,eff}(\mathcal{N}_i, T_i)}{\gamma^{(z)}_{i,eff}(m, \mathcal{N}_i)} \tag{6.9}$$

where $\overline{\gamma}_{i,eff}(\mathcal{N}_i, T_i)$ is the γ argument in (6.5) when (6.5) is set equal to $BLER_{tgt}$ for given arguments $\mathcal{N}_i$ and T_i (or using the approximation given in (6.6)). The above corresponds to the power needed for allocation.

6.2 Scheduling Ideology

Recent research, particularly for LTE systems [79–81], proposes segmenting the packet scheduling problem into a time-domain packet scheduling (TDPS) and frequency-domain packet scheduling (FDPS) problem. Such methods choose a set of users at

each scheduling interval to schedule (TDPS) and allocate them in frequency (FDPS) to maximize cell throughput (rate adaptive, RA).

In the energy limited regime, the goal is to minimize transmission energy while meeting throughput requirements. In this regime, the optimal allocation for MA differs from that of an RA problem. As a result, RA methods are not directly applicable in solving the MA problem. Further, due to limitations imposed by synchronous HARQ, the above described approaches may experience ARQ blocking since the TDPS is limited in knowledge to a single subframe in time. ARQ blocking here is defined as the inability to allocate a new TB to transmit during a given subframe due to the requirement of retransmitting a previously transmitted TB as part of the synchronous HARQ process.

A simple to implement method to eliminate ARQ blocking is to limit the number of new transmissions a station can initiate in each ARQ window. In synchronous HARQ, retransmissions are limited to exactly ARQ_w subframes following the initial transmission (where $ARQ_w = 8$ in LTE uplink) and the maximum number of times a packet can be transmitted is 4. Based on these observations, the maximum number of new transmissions that any station should initiate during any ARQ_w subframes is $ARQ_w/4 = 2$ to eliminate any ARQ blocking. Based on this, we can define an ARQ slot as the duration of time such that each UE can be allotted at most one new TB for transmission. Consequently, each ARQ slot m has a duration of 4 subframes or 1/2 of an ARQ_w (*i.e.*, $T_e = 4T_f$).

Using this segmentation approach, we can formulate the resource allocation problem by segmenting it into a two-stage formulation, namely a TDPS and block time-frequency domain packet scheduler (BTFDPS). At the beginning of scheduling epoch (ARQ slot) m, the TDPS chooses which UEs can access the channel (and with how much data, $T_i(m)$) based on the buffers and priority of each UE over ARQ slot m. The BTFDPS then allocates the UEs within the Time-Frequency grid of a ARQ slot based on channel conditions and knowledge about on-going retransmissions. The benefit of such an approach as described above is that it reduces the number of scheduling decisions needed in time as well as eliminates ARQ blocking. The drawback of this however is that it requires estimates of the channel for the overall ARQ slot.

6.2.1 Time Domain Packet Scheduling

During ARQ slot m, the TDPS determines the amount of new data to be transmitted over the channel. The TDPS decisions can be done in a number of ways. Existing works have proposed TDPS designs which demonstrate methods of exploiting information including but not limited to buffer occupancy, and quality of service requirements. The focus of this chapter is on the design of the BTFDPS and for the remainder this chapter, we assume a static TDPS where all UEs are allocated a new TB of size $T_i(m)$ during ARQ slot m.

Further, the TDPS may define a static (or dynamic) weight parameter $\alpha_i(m)$ denoting the relative priority of each station. Without loss of any generality we can assume $\sum_{i=1}^{K} \alpha_i(m) = 1, \forall m$. Weights can be designed to satisfy fairness or priority criteria. In our work, we do not focus on the design of weights and simply assume they are known.

6.2.2 Block Time-Frequency Domain Packet Scheduling

The goal of the BTFDPS is to allocate both retransmissions and new TBs within the Time-Frequency grid of an ARQ slot. For a given set of TBs from each UE $\{T_1(m), T_2(m), \ldots, T_K(m)\}$ for all m, channel state information $\boldsymbol{\gamma}$, and knowledge about TBs for retransmission, the BTFDPS applied power minimization problem can be formulated to solve for resource assignment. The resource assignment policy is comprised of the resource element assignment ($S_{i,j,n}(m)$ and $\widetilde{S}_{i,j,n}(m), \forall i, j, n, m$), power assignment ($P_{i,n}(m), \forall i, n, m$), and rate assignment ($k_{i,n}(m), \forall i, n, m$) where $n = 0, 1, 2, 3$ are the individual subframes within an ARQ slot. The quantities $S_{i,j,n}(m)$ and $\widetilde{S}_{i,j,n}(m)$ are binary indicators used to denote whether a given RB j is allocated to a given UE i for both new transmissions and retransmissions respectively, $P_{i,n}(m)$ denotes the power allocated to UE i and $b_{i,n}(m)$ denotes the spectral efficiency of the MCS chosen for UE i all during subframe n of ARQ slot m.

The design of the BTFDPS policy requires the following constraints to be satisfied:

- Minimum rate constraints (to ensure the TB is allocated)

- HARQ constraints (to ensure the retransmitted TB is allocated 8 LTE subframes following its failed transmission)
- New transmission limitation (only a single TB can be allocated per ARQ slot m)
- Contiguity constraints (to ensure RBs for a single TB are allocated contiguously in frequency)
- Allocation constraints (to ensure at most one UE can occupy a given RB during any subframe)
- BLER constraints (to ensure power is adjusted to meet $BLER_{tgt}$)

Retransmissions are handled as follows. Due to the synchronous HARQ mechanism, any transmission that was erroneous, and has not exceeded the maximum number of transmissions is rescheduled exactly 8 (ARQ_w) subframes following its original transmission. These transmissions are scheduled using the same set of RBs[2] and using the same MCS.

6.3 Optimal Allocation Formulation

Due to the dependance on previous retransmission, m, it is not possible to solve the optimal allocation for all time m. We therefore formulate the allocation problem to solve for the locally optimal allocation problem in each ARQ slot m as follows.

2. While non-adaptive HARQ limits retransmission to the same MCS, number of resource blocks and subframe, there is no limitation on which resource blocks are assigned. In our work we further limit retransmissions to the same RBs in frequency. This reduces signalling requirements and problem complexity. Retransmissions can easily be incorporated in the proposed model by considering retransmission blocks as additional users. The impact of such an assumption is discussed in the chapter summary.

6.3.1 Optimization Problem

The per-subframe power-optimal BTFDPS is formulated as follows. First, the objective function can be given as

$$\min \left(\sum_{i=1}^{K} \sum_{j=1}^{M} \sum_{n=0}^{3} \alpha_i(m)(S_{i,j,n}(m) + \widetilde{S}_{i,j,n}(m))P_{i,n}(m) \right) \tag{6.10}$$

where $\alpha_i(m)$ is the relative importance parameter of power minimization for UE i and subframe $n \in \{0, 1, 2, 3\}$ within ARQ slot m defined by the TDPS.

6.3.1.1 Rate Constraints

The amount of new information that must be transmitted during a frame is constrained as

$$\sum_{j=1}^{M} \sum_{n=0}^{3} S_{i,j,n}(m) b_{i,n}(m) \geq T_i(m), \quad \forall i \tag{6.11}$$

6.3.1.2 HARQ Constraints

The ARQ constraints ensure that resource allocation cannot occur in a subframe where a retransmission exists. This means $S_{i,j,n}(m) = 0, n \in \mathcal{S}_i(m-2), \forall i$, where $\mathcal{S}_i(m-2)$ is the set of subframes during ARQ slot $m-2$ whose transmissions were unsuccessful and have not reached the maximum number of retransmissions. Further, the number of RBs allocated to a user during a retransmission must equal the original amount of RBs assigned, *i.e.,*

$$\sum_{j=1}^{M} \widetilde{S}_{i,j,n}(m) = \sum_{j=1}^{M} (\widetilde{S}_{i,j,n}(m-2) + S_{i,j,n}(m-2)), n \in \mathcal{S}_i(m-2), \quad \forall i \tag{6.12}$$

6.3.1.3 New Transmission Limitation

In order to ensure ARQ blocking does not occur, each station is limited to a new transmission in a ARQ slot. This constraint is given as

$$\sum_{n \in \mathcal{S}_i^C(m-2)} I\left(\sum_{j=1}^{M} S_{i,j,n}(m)\right) \leq 1, \quad \forall i \tag{6.13}$$

where $I(x) = 0$ when $x = 0$ and 1 otherwise and $\mathcal{S}_i^C(m-2)$ is the complementary set of $\mathcal{S}_i(m-2)$.

6.3.1.4 Allocation and Contiguity Constraints

Additionally, localized SC-FDMA is limited to RB allocations in contiguity, this constraint can be given jointly as

$$\sum_{i=1}^{K}(S_{i,j,n}(m) + \widetilde{S}_{i,j,n}(m)) \leq 1, S_{i,j,n}(m), \widetilde{S}_{i,j,n}(m) \in \{0,1\}, \forall j,n \tag{6.14}$$

ensuring only a single user can occupy an RB during an instant of time and following from [82] as

$$S_{i,j,n}(m) - S_{i,j+1,n}(m) + S_{i,x,n}(m) \leq 1, x = j+2, \ldots, M, \forall i,j,n \in \mathcal{S}_i^C(m-2) \tag{6.15}$$

$$\widetilde{S}_{i,j,n}(m) - \widetilde{S}_{i,j+1,n}(m) + \widetilde{S}_{i,x,n}(m) \leq 1, x = j+2, \ldots, M, \forall i,j,n \in \mathcal{S}_i(m-2) \tag{6.16}$$

6.3.1.5 Power Level Allocation

The applied power allocation level is given from (6.9) and expressed as

$$P_{i,n}(m) = \frac{\overline{\gamma}_{i,eff}(\mathcal{N}_{i,n}, T_i(m))}{\gamma_{i,eff}^{(0)}(m, \mathcal{N}_{i,n})}, \forall i, n \in \mathcal{S}_i^C(m-2) \tag{6.17}$$

$$P_{i,n}(m) = \frac{\overline{\gamma}_{i,eff}(\bar{\mathcal{N}}_{i,n}, T_i(m))}{\gamma_{i,eff}^{(z_{i,n})}(m, \bar{\mathcal{N}}_{i,n})}, \forall i, n \in \mathcal{S}_i(m-2) \tag{6.18}$$

where $z_{i,n}(m)$ is the retransmission number in subframe n for UE i during ARQ slot m and $\mathcal{N}_{i,n} = \{j|S_{i,j,n}(m) = 1\}$ and $\bar{\mathcal{N}}_{i,n} = \{j|\widetilde{S}_{i,j,n}(m) = 1\}$. We note that we do not place an upper limit on the power level directly, however as shown in the next section, the goal is to minimize the average allocated power at each ARQ slot m.

Combining (6.10)-(6.18), forms the optimal BTFDPS allocation for inputs $\alpha_i(m)$, $T_i(m), \forall m$. Due to the time dependance on m and the instantaneous channel conditions, the above must be solved for each ARQ slot m online.

6.3.2 Optimization Formulation Using Binary Programming

The above problem can be formulated using a similar approach as that used in [83]. In this fashion, the contiguity constraints are exploited in a manner that reduces the binary search space.

The optimization problem is solved at each ARQ slot m. For the remainder of this section the index m is dropped however all quantities are assumed to be specific to ARQ slot m. The problem can be expressed as a general set-packing problem and formulated using binary programming as

$$\begin{aligned} \min_{\mathbf{x}} \quad & \mathbf{c}^T\mathbf{x} \\ \text{s.t.} \quad & \mathbf{A}\mathbf{x} \leq \mathbf{1}_{4M},\ \mathbf{A}_{eq}\mathbf{x} = \mathbf{1}_K, \quad x_j \in \{0,1\}, \forall j \in \mathbf{x} \end{aligned} \tag{6.19}$$

where $\mathbf{c}$ is the real-valued vector containing the weighted power of choosing a given allocation, $\mathbf{x}$ is the vector of allocation selections, $\mathbf{A}_{eq}$ is a binary equality constraint matrix of K rows and $\mathbf{A}$ is a binary inequality constraint matrix of $4M$ rows. Each non-zero entry of the solution vector $\mathbf{x}$ corresponds to selecting the corresponding column allocation in $\mathbf{A}$.

6.3.2.1 Inequality Constraints

The matrix $\mathbf{A}$ describes the set of potential RB allocations for all users. It is comprised of individual allocations given as

$$\mathbf{A} = [\mathbf{A}_1,\ \cdots,\ \mathbf{A}_K] \tag{6.20}$$

where $\mathbf{A}_i$ is a matrix containing the set of feasible allocations for UE i. Each column of $\mathbf{A}_i$ corresponds to a feasible allocation while each row corresponds to a specific resource. Within a given column, the k^{th} row corresponds to the $(k \bmod M)^{\text{th}}$ RB and subframe $n = \lfloor \frac{k-1}{M} \rfloor$ of ARQ slot m.

Each entry in $\mathbf{A}_i$ can take a value of $\{0, 1\}$. An entry of 1 if the particular resource is required by a UE for that allocation and 0 otherwise.

The set of possible allocation is determined as follows for each UE. During any ARQ slot, a UE is allocated a TB of T_i bits. For a given T_i, the maximum and minimum number of RBs can be found from Table 6.1 using b_{min} and b_{max} as

$$N_{min} = \left\lceil \frac{T_i}{132 b_{max}} \right\rceil \tag{6.21}$$

$$N_{max} = \min\left(\left\lceil \frac{T_i}{132 b_{min}} \right\rceil, M \right) \tag{6.22}$$

The effective MCS scheme is a function of the number of RBs and T_i. For each possible contiguous block of RBs of size $[N_{min}, N_{max}]$, the power level needed to maintain $BLER_{tgt}$ for all possible allocations of contiguous resource blocks is found using (6.5). For each such possible contiguous blocks above and each subframe of ARQ slot m, a column allocation in given $\mathbf{A}_i$ with corresponding transmission power in the corresponding entry of $\mathbf{c}$.

Example:

To demonstrate how to generate $\mathbf{A}_i$, consider the following example. Suppose during ARQ slot m and for UE i, $T_i(m) = 1000$ bits and there are 10 uplink resource blocks ($M = 10$). Using the set of CQIs in Table 6.1, we see that obtain 2 and 10 as the minimum and maximum unique quantity of RBs that satisfy the above. Therefore we let all eligible RB allocations fall within 2 and 10 RBs inclusive $[2, 10]$ (we denote this set $\mathcal{E}$) with spectral efficiencies given as solving (6.1) set equal to $T_i(m)$.

Suppose we consider $M'(b_i) = 6$. It can easily be shown that for $M = 10$ (to simplify the example), there are 5 $(M - M'(b_i) + 1 = 10 - 6 + 1)$ ways to allocate 6 RBs contiguous in frequency. As a result, we define a submatrix $\bar{\mathbf{A}}_i^{(6)}$ computed as

the 5 possible allocations or as

$$\bar{\mathbf{A}}_i^{(6)} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{6.23}$$

Next, we define ${}^{(n)}\bar{\mathbf{A}}_i^{(6)}$ as the eligible allocations of 6 RBs for subframe n. It is given as

$${}^{(n)}\bar{\mathbf{A}}_i^{(6)} = \begin{cases} \bar{\mathbf{A}}_i^{(6)}, & \text{if User } i \text{ has no ReTx in } n \\ \emptyset, & \text{otherwise} \end{cases} \tag{6.24}$$

where here $\emptyset$ is used to denote a 0×0 (empty) matrix. From this we obtain $\mathbf{A}^{(6)}$ as

$$\mathbf{A}_i^{(6)} = \begin{bmatrix} {}^{(0)}\bar{\mathbf{A}}_i^{(6)} & {}^{(1)}\mathbf{0} & {}^{(2)}\mathbf{0} & {}^{(3)}\mathbf{0} \\ {}^{(0)}\mathbf{0} & {}^{(1)}\bar{\mathbf{A}}_i^{(6)} & {}^{(2)}\mathbf{0} & {}^{(3)}\mathbf{0} \\ {}^{(0)}\mathbf{0} & {}^{(1)}\mathbf{0} & {}^{(2)}\bar{\mathbf{A}}_i^{(6)} & {}^{(3)}\mathbf{0} \\ {}^{(0)}\mathbf{0} & {}^{(1)}\mathbf{0} & {}^{(2)}\mathbf{0} & {}^{(3)}\bar{\mathbf{A}}_i^{(6)} \end{bmatrix} \tag{6.25}$$

where ${}^{(n)}\mathbf{0}$ is a matrix of zeros identical to size of ${}^{(n)}\mathbf{A}_i^{(6)}$. The above corresponds to the feasibility of allocating any user in any subframe of the ARQ slot in which it does not have an ongoing retransmission. Finally we obtain $\mathbf{A}_i$ as

$$\mathbf{A}_i = \left[\mathbf{A}_i^{(2)}, \mathbf{A}_i^{(3)}, \mathbf{A}_i^{(4)}, \mathbf{A}_i^{(5)}, \mathbf{A}_i^{(6)}, \mathbf{A}_i^{(7)}, \mathbf{A}_i^{(8)}, \mathbf{A}_i^{(9)}, \mathbf{A}_i^{(10)}\right] \tag{6.26}$$

where the above is found by concatenating each submatrix obtained from each entry of $\mathcal{E}$.

6.3.2.2 Equality Constraints

The equality matrix $\mathbf{A}_{eq}$ is simply a matrix of K rows constraining the number of selected allocations such that each UE is only allotted one allocation selection from their matrix $\mathbf{A}_i$. This is given as

$$\mathbf{A}_{eq} = \begin{bmatrix} \mathbf{1}_{C_1}^T & \cdots & \mathbf{0}_{C_K}^T \\ \vdots & \ddots & \vdots \\ \mathbf{0}_{C_1}^T & \cdots & \mathbf{1}_{C_K}^T \end{bmatrix} \tag{6.27}$$

where C_i is the number of columns in $\mathbf{A}_i$ and $\mathbf{1}_x$ and $\mathbf{0}_x$ are column vectors of length x.

6.3.2.3 Cost Function

The objective function vector $\mathbf{c}^T = [\mathbf{c}_1^T, \ldots, \mathbf{c}_K^T]$ is simply the cost of choosing the corresponding allocation for each $\mathbf{c}_i$. From (6.19) we see the cost is simply the weighted power of choosing an allocation. Since the power is directly a function of the set of resources, the channel, and the TB size, individual entries of $\mathbf{c}_i$ can be then be given as

$$c_{i:j_i} = \alpha_i P_i(\mathcal{N}_i(j_i), T_i), \quad j_i = 1, 2, \ldots, C_i \tag{6.28}$$

where the function $P_i(\cdot)$ is given in (6.9), α_i is the priority weight of UE i and where $\mathcal{N}_i(j_i) = \{x | a_{i:x,j_i} = 1, x = 1, 2, \ldots, M\}$. The quantity $a_{i:x,j_i}$ denotes the $\{x, j_i\}$ entry in $\mathbf{A}_i$ and j_i is the j_i^{th} column of $\mathbf{A}_i$.

The above problem is solved at each ARQ slot m online for given $T_i(m), \forall i$ [3]. While the above will yield the optimal solution at each decision epoch, and is less computationally intensive than the OFDM MA method with additional contiguity constraints as the search space is reduced pre-computation. The resulting framework is still computationally prohibitive for online operation. In the next section, we propose two methods to reduce the problem complexity.

3. $T_i(m)$ can be a constant or vary in time

6.4 Sub-Optimal Resource Allocation Schemes

Due to the rather large search space above, we propose two sub-optimal methods for the above approach. The first approach is a subset reduction technique which is herein referred to as the Best-N scheme. In this method, we exploit the potential of choosing a subset of the best allocations (in terms of minimizing transmission power) for each user and allocating to satisfy this set. The second method, herein denoted as the iterative allocation method, allocates time-frequency resources to all users to maximize the *power level gain* at each iteration. Both methods are described in the following sections.

6.4.1 Method 1 - Best-N Subset Reduction

The Best-N method provides for a varying degree of complexity and operates as follows. By including a limited number (N) of the most energy efficient allocations for a UE determined based on channel measurements (a subset of matrix $\mathbf{A}_i$), the optimization framework can consider a reduced search space. The search space increases as N increases, and in the limiting case when N is large, this method approaches the previously described optimal allocation at each ARQ slot m. The algorithm operates as follows. At the beginning of each ARQ slot m determine the subset of best allocations available for transmission as described in below. Each allocation corresponds to an allocation in the matrix $\mathbf{A}_i$ described previously.

The Best-N allocations represent the N allocations (set of RBs, MCS mode, and power allocated) that are the most energy efficient for the UE to use for transmission. This is essentially a pre-computed subset matrix of the previously described $\mathbf{A}_i$. We denote this new constraint matrix as $\mathbf{B}_i$.

Once efficient allocations have been computed, the revised set-packing problem and can be formulated using binary programming as

$$\begin{aligned} \min_{\mathbf{x}} \quad & \mathbf{d}^T\mathbf{x} \\ \text{s.t.} \quad & \mathbf{B}\mathbf{x} \leq \mathbf{1}_{4M},\ \mathbf{B}_{eq}\mathbf{x} = \mathbf{1}_K, \quad x_j \in \{0,1\}, \forall j \in \mathbf{x} \end{aligned} \tag{6.29}$$

where $\mathbf{d}$ is a real-valued vector, $\mathbf{x}$ is the vector of allocation selections, $\mathbf{B}_{eq}$ is the equality constraint matrix of K rows and $\mathbf{B}$ is the inequality constraint matrix of $4M$ rows. Each non-zero entry of the solution vector $\mathbf{x}$ corresponds to selecting the corresponding column allocation in $\mathbf{B}$.

The matrix $\mathbf{B}$ is found as

$$\mathbf{B} = [\mathbf{B}_1, \ \cdots, \ \mathbf{B}_K] \tag{6.30}$$

where entries $\mathbf{B}_i$ are given as

$$\mathbf{B}_i = \begin{bmatrix} {}^{(0)}\bar{\mathbf{B}}_i & {}^{(1)}\mathbf{0} & {}^{(2)}\mathbf{0} & {}^{(3)}\mathbf{0} \\ {}^{(0)}\mathbf{0} & {}^{(1)}\bar{\mathbf{B}}_i & {}^{(2)}\mathbf{0} & {}^{(3)}\mathbf{0} \\ {}^{(0)}\mathbf{0} & {}^{(1)}\mathbf{0} & {}^{(2)}\bar{\mathbf{B}}_i & {}^{(3)}\mathbf{0} \\ {}^{(0)}\mathbf{0} & {}^{(1)}\mathbf{0} & {}^{(2)}\mathbf{0} & {}^{(3)}\bar{\mathbf{B}}_i \end{bmatrix} \tag{6.31}$$

where $\mathbf{0}$ is a matrix of zeros identical to size of ${}^{(n)}\bar{\mathbf{B}}_i$ and ${}^{(n)}\bar{\mathbf{B}}_i$ is

$${}^{(n)}\bar{\mathbf{B}}_i = \begin{cases} \bar{\mathbf{B}}_i, & \text{if User } i \text{ has no ReTx in } n \\ \emptyset, & \text{otherwise} \end{cases} \tag{6.32}$$

The matrix $\bar{\mathbf{B}}_i$ is an $M \times N$ matrix containing the set of Best-N allocations. The matrix $\bar{\mathbf{B}}_i$ is obtained as follows. Let $\mathcal{N}(i,n)$ be the set of RB allocations for the n^{th} best[4] allocation of i. Then we have entries of $\bar{\mathbf{B}}_i$ as

$$b_{i:n,j} = \begin{cases} 1, & j \in \mathcal{N}(i,n) \\ 0, & \text{otherwise} \end{cases}, \quad j = 1, 2, \ldots, M \tag{6.33}$$

4. The n^{th} best allocation refers to the allocation that achieves the n^{th} smallest required power allocation subject to any additional constraints.

Consequently, the matrix $\mathbf{B}_{eq}$ is given as

$$\mathbf{B}_{eq} = \begin{bmatrix} \mathbf{1}^T_{\bar{C}_1} & \cdots & \mathbf{0}^T_{\bar{C}_K} \\ \vdots & \ddots & \vdots \\ \mathbf{0}^T_{\bar{C}_1} & \cdots & \mathbf{1}^T_{\bar{C}_K} \end{bmatrix} \tag{6.34}$$

where $\bar{C}_i$ is used to denote the number of columns in $\mathbf{B}_i$. The vector $\mathbf{d}$ is obtained identical to $\mathbf{c}$ for entries corresponding to allocations in $\mathbf{B}$.

In rare cases the solution may be infeasible during a given scheduling epoch as we do not exhaustively allow selection of every possible allocation. To minimize the probability of this occurrence, the following steps can be taken:

- Employ appropriate admission control design in the TDPS.
- Limit the maximum number of resource blocks that can be allocated to a single user as a function of available resources and required data rate.
- Enforce overlap restrictions on the Best-N selection scheme such that no two possible allocations for a single UE may share any resource blocks (employed in this implementation).
- Increment N and resolve the problem if the solution is found infeasible (employed in this implementation).

There are no guarantees that the Best-N method will guarantee a solution without large increases in N, if the system is highly loaded (size of TBs and number of users is large). As a result, it is not an ideal approach to the allocation problem. In the next section, we propose an additional method that iteratively allocates RBs to each user.

6.4.2 Method 2 - Iterative Allocation

The second suboptimal method tries to iteratively allocate resources to all UEs. At each iteration, resources are allocated to the user to maximize the *power level gain*. This method is described mathematically in Algorithm 1. The function $\mathfrak{F}(x, i)$ returns the subframe index corresponding to index x for UE i and $\mathfrak{G}(N, \mathcal{N}_i, \mathcal{Z}_i)$ returns a set

$\mathcal{N}_i$ of subsets $\mathcal{N}_{i,x}$ containing all unique contiguous RBs of size N in all subframes $n \notin \mathcal{Z}_i$ (subframes which are not allocated for retransmission for UE i). The number of subsets in $\mathcal{N}_i$ is $|\mathcal{N}_i|x$ and the size of each subset $\mathcal{N}_{i,x}$ is $|\mathcal{N}_{i,x}| = N$.

The proposed algorithm operates as follows. The initialization of the algorithm is described by lines 1-12. The iterative allocation portion of the algorithm (described in lines 13-51 inclusive) is divided into two major components: the power level gain computation (lines 13-33) and the allocation stage (lines 37-50).

6.4.2.1 Initialization Procedure

The initialization stage is described through lines 1 to 12. As with the previously describe solutions, retransmissions are allocated first to determine the residual resources available for new transmissions. Additionally, the minimum and maximum amount of number of resource blocks for each UE i is determined.

6.4.2.2 Power Level Gain

The *power level gain* (PLG) stage operates as follows. For users who have been already allocated resources, the PLG of a resource is calculated as the difference in power allocated if the new resource is added to the current resource allocation and the currently allocated resources (line 17). The newly added resource is constrained

Algorithm 1 Iterative Resource Allocation: Initialization

1: $\mathcal{K} = \{1, 2, \ldots, K\}$
2: *Allocate ReTx with required power*
3: $\mathcal{N}_n^{(r)} \leftarrow$ *Set of RBs in n with ReTx*, $n = 0, 1, 2, 3$
4: $\mathcal{Z}_i \leftarrow$ *Set of subframes with ReTx for UE i*, $n = 0, 1, 2, 3 \ \ \forall i \in \mathcal{K}$
5: $\mathbf{n}(i) = -1, \ \ \forall i \in \mathcal{K}$
6: $\mathcal{N}_n^{(a)} = \{1, 2, \ldots, M\} \setminus \mathcal{N}_n^{(r)}, \ \ n = 0, 1, 2, 3$
7: $\mathcal{N}_i = \emptyset, \quad \forall i \in \mathcal{K}$
8: **for** $i \in \mathcal{K}$ **do**
9: $\quad \mathcal{N}_i^{(f)} = \bigcup_{n=0, n \neq \mathcal{Z}_i}^{3} \mathcal{N}_n^{(a)}$
10: $\quad N_{min,i} = \left\lceil \frac{T_i(m)}{132 b_{max}} \right\rceil$
11: $\quad N_{max,i} = \min\left(\left\lceil \frac{T_i(m)}{132 b_{min}} \right\rceil, M \right)$
12: **end for**

Algorithm 1 Iterative Resource Allocation: Iterative Component

13: **while** $\max(PLG_i) > 0$ && $\bigcup_{n=0}^{3} \mathcal{N}_n^{(a)} \neq \emptyset$ **do**
14: **for** $i \in \mathcal{K}$ **do**
15: **if** $\mathcal{N}_i \neq \emptyset$ && $|\mathcal{N}_i| < N_{max,i}$ **then**
16: **for** $j \in \mathcal{N}_i^{(f)} \cap \mathcal{N}_{\mathbf{n}(i)}^{(a)}$ **do**
17: $\Delta p_{i,j} = \alpha_i (P(\mathcal{N}_i, T_i, \boldsymbol{\gamma}) - P(\mathcal{N}_i \cup j, T_i, \boldsymbol{\gamma}))$
18: **end for**
19: $PLG_i = \max(\{\Delta p_{i,j} | j \in \mathcal{N}_i^{(f)} \cap \mathcal{N}_{\mathbf{n}(i)}^{(a)}\})$
20: **else**
21: $\mathcal{N}_i^{(c)} = \mathfrak{G}(N_{min,i}, \{\mathcal{N}_n^{(a)} | \forall n\})$
22: **if** $|\mathcal{N}_i^{(c)}|_x == 1$ **then**
23: $i^* = i$
24: $x^* = 0$
25: *Go To Line* 38
26: **end if**
27: **for** $x = 0; x < |\mathcal{N}_i^{(c)}|_x$ **do**
28: $p_{i,x} = \alpha_i P(\mathcal{N}_{i,x}^{(c)}, T_i, \boldsymbol{\gamma})$
29: **end for**
30: $\mathcal{N}^{(B)} = \{0, 1, \ldots, |\mathcal{N}_i^{(c)}|_x - 1\}$
31: $PLG_i = \min(\{p_{i,x}, x \in \mathcal{N}^{(B)} \setminus \underset{x^* \in \mathcal{N}^{(B)}}{\arg\min}\ (p_{i,x^*})\}) - \min(\{p_{i,x}, \forall x \in \mathcal{N}^{(B)}\})$
32: **end if**
33: **end for**
34: **if** $\max(PLG_i) < 0$ **then**
35: **break**
36: **end if**
37: $i^* = \underset{i}{\arg\max}(PLG_i)$
38: **if** $\mathbf{n}(i^*) == -1$ **then**
39: $x^* = \underset{x}{\arg\max}(\{p_{i,x} | i = i^*\})$
40: $\mathcal{N}_{i^*} = \mathcal{N}_{i^*,x^*}^{(c)}$
41: $\mathbf{n}(i^*) = \mathfrak{F}(x^*, i^*)$
42: $n^* = \mathbf{n}(i^*)$
43: $\mathcal{N}_{n^*}^{(a)} = \mathcal{N}_{n^*}^{(a)} \setminus \mathcal{N}_{i^*,x^*}^{(c)}$
44: **else**
45: $n^* = \mathbf{n}(i^*)$
46: $j^* = \underset{j}{\arg\max}(\{\Delta p_{i,j} | i = i^*\})$
47: $\mathcal{N}_{i^*} = \mathcal{N}_{i^*} \cup j^*$
48: $\mathcal{N}_{n^*}^{(a)} = \mathcal{N}_{n^*}^{(a)} \setminus j^*$
49: **end if**
50: $\mathcal{N}_{i^*}^{(f)} = \{\min(\mathcal{N}_{i^*}) - 1, \max(\mathcal{N}_{i^*}) + 1\} \cap \mathcal{N}_{n^*}^{(a)}$
51: **end while**

to those resources within a users feasible allocation set ($\mathcal{N}_i^{(f)}$). Alternatively, for any user who has yet to be allocated a resource, the PLG of that user is measured as the difference between the power allocation required if that best available resource is allocated and that of the second best available resource (line 31).

If at any instant of time, a user has only one eligible allocation, the user is allocated that resource (described in lines 22-25).

6.4.2.3 Resource Allocation Stage

The user with the maximum PLG at any iteration is allocated the corresponding resource (line 38-50). For users being allocated an initial resource, this constrains the feasible resources for allocation within that given subframe n.

The overall allocation continues until the maximum PLG is negative or there is no additional resources for transmission.

6.5 Numerical Evaluation

Simulations results are provided measuring the power and computation timing of the optimal, Best-N and iterative scheme. In addition, we measure the impact of static scheduling of retransmissions. Default simulation parameters are given in Table 6.4.

6.5.1 Least-Squares Fit Function

The justification behind use of the pre-described fit function is shown in Figure 6.3. Here we see in this comparison for $2, 4$, and 8 RBs with $BLER_{tgt} = 10\%$. The

Table 6.4: Simulation Parameters

Parameter	Value	Parameter	Value
γ_0	10 dB	T_f	1ms
$\alpha_i(m)$	1 ($\forall i$)	T_e	4ms
N_{ctrl} (symbols)	3	N (Best-Ndepth)	3
N_{sym} (symbols)	14	$\bar{T}_i$	250 bits
Number of Subframes	20000	$BLER_{tgt}$	10%

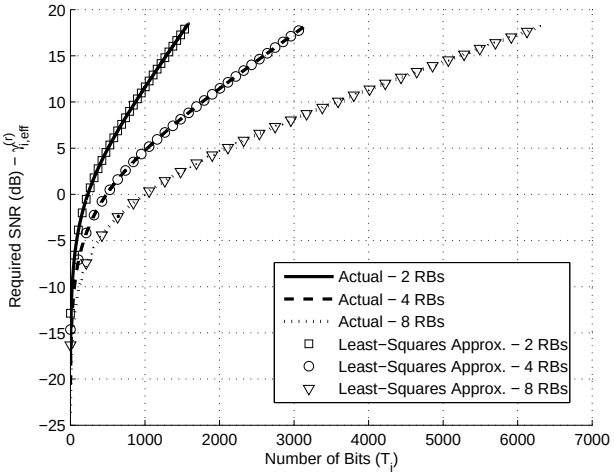


Figure 6.3: Least-Squares Approximation Accuracy

least-squares approximation is shown to hold tightly to the actual BLER function; providing for a more tractable computation of the required SNR level and justifying its use as a suitable alternative in computation of the required SNR.

6.5.2 Retransmission Power - Static Scheduling

Worth noting is the importance of the impact of static scheduling on retransmissions. While using the above framework, packet error rates are 10%, the overall power allocated for retransmission packets accounts for over 50% of allocated power. As a result, in some simulations, the Best-N method requires less overall power to be allocated than the optimal method as a result of power expended for retransmissions. Overall, the iterative method tends to perform worse in this area in all scenarios as in general, the average number of resource blocks assigned to any user is lower. This is particularly detrimental if a user experiences a low quality channel in those RB(s) as the channel is not averaged over a larger set of resources.

This overall result highlights the importance of re-allocating TBs for retransmission within a given ARQ slot. To accomplish this, the previously described mechanisms could be modified as follows. For both the optimal and Best-N scheme, each user with a retransmission can be considered as an additional user to allocate. Unlike

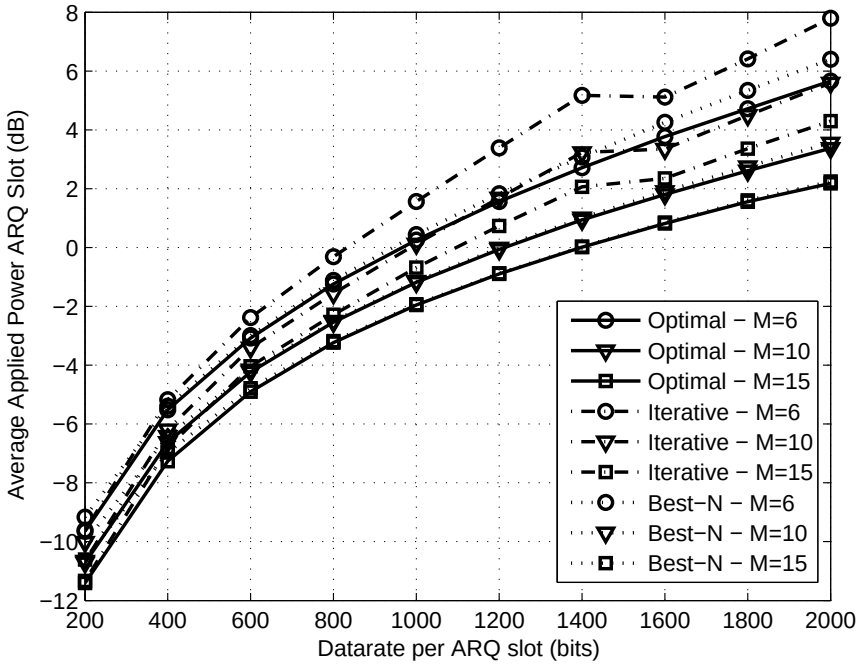
new transmissions, these transmission have a fixed number of RBs, in addition are limited to a single subframe. As a result, there is a relatively small complexity increase for these additional users. For the iterative mechanism, retransmissions can be incorporated by considering them as additional users. In this case, $N_{min} = N_{max} = |\mathcal{N}_i|$ where here $\mathcal{N}_i$ denotes the set of resource blocks for the initial transmission of a given TB (replacing lines 10 and 11 in Algorithm 1).

6.5.3 Optimal Power Expenditure Gap

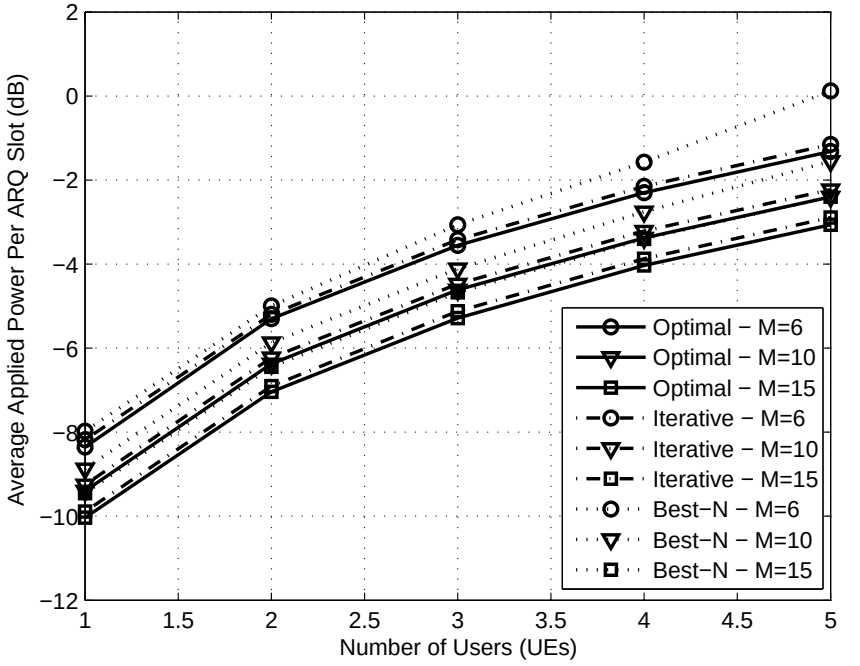
The average power expenditure per ARQ slot for the initial transmission of a TB is shown in Figure 6.4. Here we observe the following. In general the amount of power expended increases with both datarate and the number of UEs. For lower datarates, the iterative mechanism outperforms the Best-N scheme, however as the per ARQ slot datarate increases, Best-N performance surpasses that of the iterative method. At approximately $T_i = 1400$ bits, the iterative method experiences an uncharacteristic trend. This is a result of an increase to N_{min} at that datarate, forcing the iterative scheme to maximize *power level gain* of more than one RB for the initial allocation. In this way, the scheduler is better able to allocate a block of RBs. The effect of the number of UEs on the power between the two suboptimal methods is negligible as it largely depends on the datarate. Both suboptimal methods obtain near optimal per ARQ slot performance.

6.5.4 Complexity Comparison

The system complexity (measured in relative computation time) is shown in Figure 6.5. This is measured as the percentage increase in computation time taken compared to the base case (iterative method and $M = 6$ for both figures and $T_i = 200$ and $K = 1$ for Figures 6.5(a) and 6.5(b) respectively). As expected, the iterative method has the lowest complexity of all schemes. Further, the number of UEs and/or RBs results in a large increase in complexity, while the datarate is shown to have negligible impact.

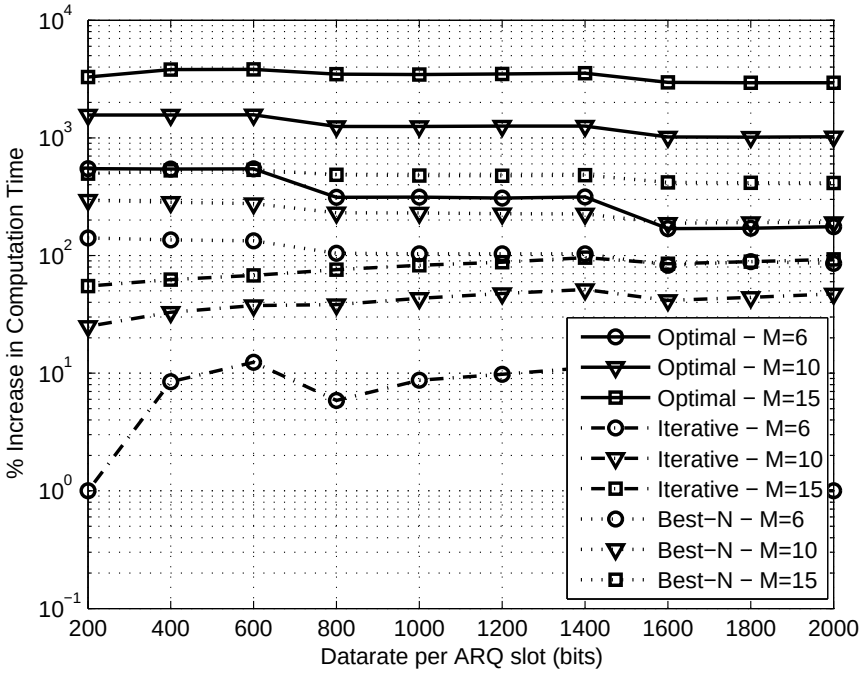


(a) versus Datarate (K=1)

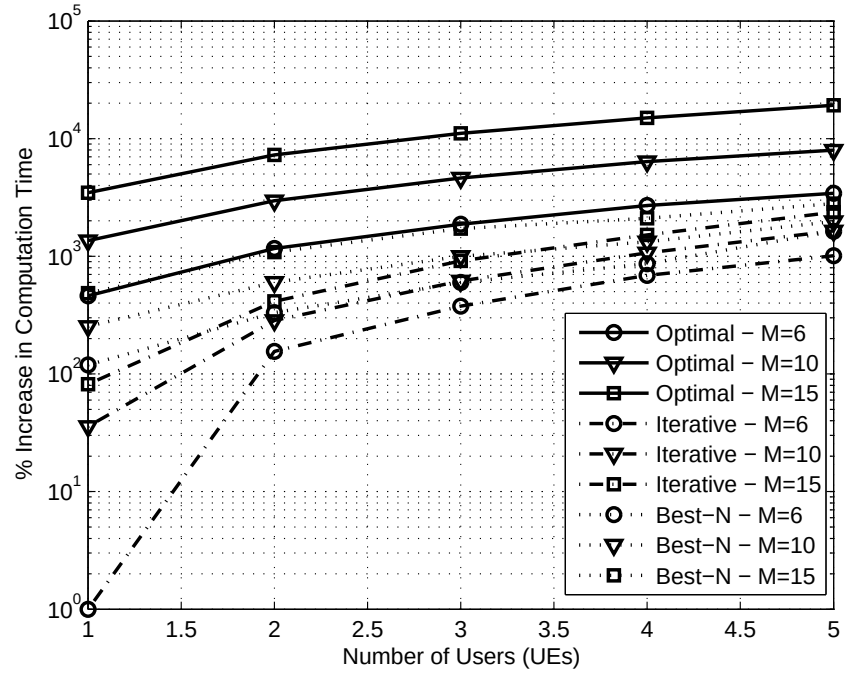


(b) versus Number of UEs

Figure 6.4: Average Applied power for Initial Transmission



(a) versus Datarate (K=1)



(b) versus Number of UEs

Figure 6.5: Computation Time Comparison

6.6 Chapter Summary

In this chapter we proposed a framework for energy efficient resource allocation in the SC-FDMA multi-user uplink. Firstly, we proposed an alternative method of selecting an appropriate scheduling epoch based on the impact of synchronous HARQ. By utilizing the proposed method, we can reduce the number of allocation procedures in time and ensure users can always initiate a new transmission during any frame (*i.e..*, do not experience ARQ blocking).

Secondly, we proposed two sub-optimal power efficient resource allocation methods. Both methods were compared to the optimal method in terms of complexity and power efficiency. We found that the sub-optimal methods closely obtain the power efficiency of the optimal allocation with reduced complexity. Further, we found that the efficiency of the power allocation scheme is dramatically reduced when static scheduling is employed for retransmissions.

One of the large drawbacks in this chapter, is the assumption on the availability of channel estimates for the duration of an ARQ slot. While this may be the case in a slow fading channel, this assumption will not hold in a fast fading environment. In future work, this issue should be addressed. The concept of an ARQ slot can be used to eliminate ARQ blocking, however allocation within the ARQ slot to individual subframes for users can be based on estimates of the evolution of the channel. Details of this, and other related open problems are discussed under future work in Chapter 8.

Chapter 7
Energy Efficient QoS Constrained Scheduler for SC-FDMA Uplink

In previous chapters we have independently looked at low complexity scheduling methods for both SC-FDMA and multiple users/streams with QSI. In this chapter, we introduce and demonstrate a method to extend this work to incorporate our energy efficiency multiple stream scheduling framework in the uplink SC-FDMA scenario. The remainder of this chapter is divided as follows. In Section 7.1 we overview the details of the employed uplink system model including the channel and scheduling models and in Section 7.2 we describe the scheduling ideology. In Section 7.3 simulation results are provided while in Section 7.4, conclusions are drawn on this work.

7.1 System Model

The system model is shown in Figure 7.1 and similar to the model presented in Chapter 6, however we do not consider the HARQ process. For clarity to the reader we briefly describe it again here.

We assume that there are K users (denoted as UEs) within a single cell, communicating with a single base station (denoted as an eNB). Since we are concerned with resource allocation within a single cell, for the purpose of our work, it is assumed that intercell interference is negligible. The cell spectrum is divided into N_{sub} subcarriers which are grouped into M resource blocks. Each resource block (RB) is comprised of 12 equivalent subcarriers. Without loss of generality we assume there is an Integer number (M) of RBs available for allocation. The system is assumed to be operating in FDD mode.

There are N_{sym} symbols per subcarrier in a given subframe where the exact number of subcarriers depends on the uplink configuration. The physical uplink shared channel (PUSCH) is used for transmission of uplink data and comprises a portion of symbols along with other controls channels. For the purpose of our work, it is assumed the PUSCH occupies $N_{sym} - N_{ctrl}$ symbols per subcarrier, per subframe where N_{ctrl} is the number of symbols used for all other physical channels and signalling.

Each UE receives uplink traffic from upper layers of their protocol stack destined for transmission to the eNB. Each UE's traffic has associated QoS parameters $\{D_i, L_i, \overline{\lambda}_i, B_i, \delta_i\}$ which denotes the maximum tolerable average delay, SDU length, average arrival rate, buffer size at the radio link control (RLC) layer and maximum tolerable packet loss rate respectively for that stream. Each stream may represent a broad service class (such as voice over IP or video) or a particular application-layer stream being used at the time. Each incoming stream is stored in a finite-length first-in, first-out (FIFO) buffer where incoming packets are dropped when the buffer is full.

The scheduling horizon is divided into subframes consisting of 2 LTE time slots (of duration 1ms). During each subframe m, users can transmit up to T_i bits of data as determined by the eNB. On average each user will receive a service rate of μ_i packets per second.

During each frame, each of the K users can transmit a single transport block of $T_i(m)$ bits. For now, it is assumed that eligible transport block sizes are an integer number of SDUs plus a header (*i.e.,* no SDU segmentation is required by the RLC). In this case, the scheduling objective for the MAC is to determine either a priori or online, a way to allocate each user a quantity of data to transmit $T_i(m)$ for all subframes m to meet the individual loss, delay and throughput requirements of all users and while minimizing the average weighted energy expenditure. The MAC layer scheduling decision is as follows. For each user, design the set of scheduling policy decisions to minimize the long-term average allocated transmission power.

The MAC allocation component is formulated as a constrained Markov decision process where the system state of user i is denoted by its buffer level and the action

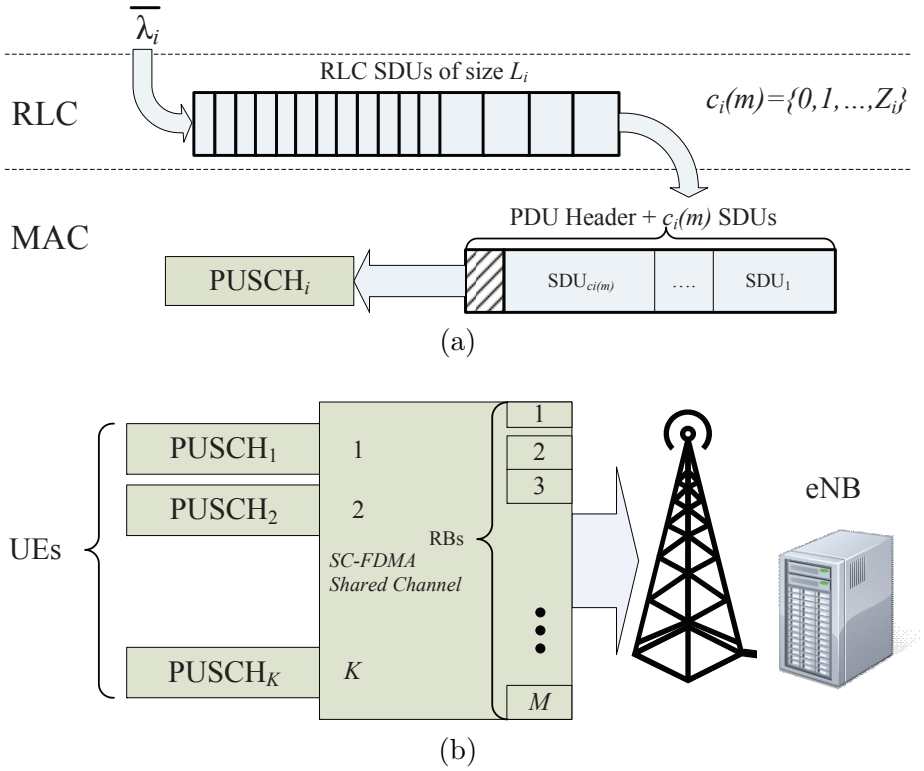


Figure 7.1: SC-FDMA Uplink System Model: (a) Each UE, (b) Overall System.

space describes the number of packets that can be transmitted during a subframe subject to a randomized policy. The solution to the problem for any scheduling policy $\mathbf{\Omega}$ is a random policy which described by the distribution $\theta_i(c_i, u_i|\mathbf{\Omega})$ which denotes the probability of choosing the action c_i as the number packets for transmission given that queue i is in state u_i. The aforementioned policy is derived for all users. The goal of the optimization formulation is to find $\theta_i(c_i, u_i|\mathbf{\Omega})$ for all c_i, u_i, as well as i that minimizes the average applied transmission power. The resultant policy is coupled by $\mathbf{\Omega}$ which defines the scheduling actions for each queue i and each queue state $u_i \in \mathcal{U}_i$. Using this policy, in each subframe m, $T_i(m)$ bits are chosen for transmission from UE i (where $T_i(m) = L_i c_i(m) + L_{hdr}$) and $c_i(m)$ is c_i chosen randomly at time m subject with probability defined by $\theta_i(c_i, u_i|\mathbf{\Omega})$.

For anytime m, $\{c_i(m)|0 \leq c_i(m) \leq Z_i, \forall i\}$ denotes the joint action space taken for all UEs. The set of all feasible joint action spaces is $\mathcal{C}$ (known a priori). In order to solve for the aforementioned policy $\mathbf{\Omega}$ to minimize transmission power, one must

first determine the power cost of taking each joint action $\mathbf{c} \in \mathcal{C}$. This is handled by the physical (PHY) layer component.

Firstly, we clarify the following assumptions

- The CSI matrix corresponding to the channel between each UE and the eNB over all RBs is available at the eNB error free.
- The eNB feedback channel informs UEs in advance of the resource blocks and quantity of data for transmission for a user during any uplink subframe.
- The eNB has knowledge about buffer occupancy levels and QoS parameters of each UE.

7.1.1 Finite Transport Block Sizes

In a packet based transmission system, due to physical limitations, each user can only service a finite number of packets from the queue. Assuming no packet segmentation at the physical layer before transmission, we can denote Z_i as the maximum number of packets (maximum value of c_i, with rationale described in Chapter 3) that can be serviced during any time subframe m by UE i where $c_i \in \mathcal{C}_i = \{0, 1 \ldots, Z_i\}$. As a result, the eligible size of each UE's transport block sizes is also finite and given as $\mathcal{T}_i = \{0, L_i + L_{hdr}, 2L_i + L_{hdr}, \ldots, Z_i L_i + L_{hdr}\}$ where L_i is the packet length in bits and L_{hdr} is the header size. During each subframe, the eNB allocates to user i an uplink slot of $T_i(m) \in \mathcal{T}_i$ bits.

7.1.2 Channel State Information (CSI)

Channel state information is assumed available at the eNB for the next scheduling time-frame. We assume this information is available error-free. The channel is modelled as block fading where the channel is static for the duration of a subframe and independent from subframe to subframe. The channel experienced from user to user is assumed independent. For now, the channel transfer function for each RB is also assumed to be independent of adjacent RBs and each channel follows the Rayleigh SNR distribution given as

$$p(\gamma) = \frac{1}{\gamma_0} \exp\left(-\frac{\gamma}{\gamma_0}\right) \tag{7.1}$$

where $\gamma_{i,k}(m)$ will be used to denote the uplink channel of user i over RB k in subframe m.

7.1.3 Queue Evolution

From slot m to slot $m+1$ the evolution of the RLC queue of each user evolves according to

$$u_i(m) = \min\{B_i, \max\{0, u_i(m-1) - c_i(m)\} + A_i(m)\} \tag{7.2}$$

where $u_i(m)$ is the number of packets in queue i at the end of subframe m, $A_i(m)$ is the number of packets arriving during subframe m to the queue and $c_i(m)$ is the number of packets taken from queue i during subframe m.

7.2 Scheduler Formulation

The scheduler formulation is defined through a scheduling policy $\mathbf{\Omega}$. Let $\theta_i(c_i, u_i|\mathbf{\Omega})$ be a steady-state distribution function that exists for a particular policy $\mathbf{\Omega}$ which denotes the probability of being in state u_i and transmitting c_i packets during frame n. The scheduling policy $\mathbf{\Omega}$ is obtained through application of constraints on average delay and MAC layer throughput which are given as follows. This approach was described in Chapter 3 where the major contributions here is in generation of the cost function used in conjunction with the MAC layer constraints in generation of the scheduling policy.

7.2.1 PHY Layer Transmission Constraints

The uplink physical layer employs SC-FDMA. Resources during any given subframe must be allocated contiguously in frequency, and only a single contiguous transport block can be allocated per subframe.

The PHY allocation is done employing an iterative, near-optimal allocation technique using Algorithm 2 or using an optimal allocation approach as described in

Appendix D. In Appendix D, it is shown that this method performs near-optimal subcarrier and power allocation when the number of users scheduled per time frame to less than half the number of the RBs (i.e., $2K \leq M$) and under the TB size of interest using the pre-described channel conditions. We assume that the number of users is less than half the number of RBs and therefore justify its use. For a given set of MAC layer allocations $\{T_i(m), \forall i\}$, the physical layer allocates power and RBs to each user. The selected power level is given as the ratio of the required SNR to the measured SNR for a given target error rate. We utilize the block outage probability from [10] to model the block error rate (BLER) of coded transmissions[1]. This is a function of the number of resource blocks (total number of symbols) and the data rate. Given a target BLER, a data rate T_i (measured in bits per scheduling frame) and a set of resource blocks $\mathcal{N}_i$, we know that

$$BLER(\gamma_{i,eff}^{(r)}, \mathcal{N}_i, T_i) \approx Q\left(\frac{\log(1+\gamma_{i,eff}^{(r)}) - \frac{\log(2)T_i}{\eta(\mathcal{N}_i)}}{\sqrt{\frac{2}{\eta(\mathcal{N}_i)}\frac{\gamma_{i,eff}^{(r)}}{1+\gamma_{i,eff}^{(r)}}}}\right) \tag{7.3}$$

where $\gamma_{i,eff}^{(r)}$ is the SNR level required for a given BLER.

In order to use the above for margin adaptive resource allocation, we must solve for $\gamma_{i,eff}^{(r)}$ as a function of T_i and $\mathcal{N}_i$. From this, one can obtain the required allocation power from the SNR gap between the measured effective SNR and the required SNR. Due to the monotonicity of the Q-function arguments, the above can be solved efficiently using bisection techniques. Alternatively, a more computationally efficient method is to obtain a least squares approximation to the above as a function of data rate, target BLER and the number of RBs allocated (similar to the approach in [20]). We found the following fitting function closes approximates the SNR as a

1. While here we utilize BLOP to model the error rate of coded transmissions, extensions are trivial given measurements of BLER performance obtained via proper training and calibration or in cases where analytical expressions are obtainable

function of data rate

$$\gamma_{i,eff}^{(r)} \approx a_x \exp(b_x T_i) - \gamma_{0,x} \tag{7.4}$$

where $x = |\mathcal{N}_i|$ and validation of this approximation is given in Section 7.3.1.

Using the above, the required applied power is given as

$$P(\mathcal{N}_i, T_i, \boldsymbol{\gamma}) = \frac{\gamma_{i,eff}^{(r)}}{\gamma_{i,eff}^{(m)}} = \frac{a_{|\mathcal{N}_i|} \exp(b_{|\mathcal{N}_i|} T_i) - \gamma_{0,|\mathcal{N}_i|}}{\frac{1}{|\mathcal{N}_i|^2} \sum_{k \in \mathcal{N}_i} \gamma_{i,k}} \tag{7.5}$$

where $\mathcal{N}_i$ is the set of RBs that we allocate to a UE and T_i is the number of bits for transmission. Parameters a_x, b_x, and $\gamma_{0,x}$ are given from the least-squares approximation derived in Chapter 6 depends on the target block error rate of the channel ($BLER_{tgt}$).

7.2.2 MAC Layer Constraints

The primary MAC layer constraints are on throughput and delay. These are measured as follows.

7.2.2.1 MAC Throughput

Throughput is measured as the amount of goodput over the channel. Assuming each transmission experiences a block error rate of $BLER_{tgt}$ the average throughput is given by[2]

$$Throughput = \mathbb{E}_m[T_i(m)](1 - BLER_{tgt}) \tag{7.6}$$

Further, we note that dropping probability of a given queue is related to the service rate as

$$P_{drop,i} = 1 - \frac{\mathbb{E}_m[T_i(m)]}{\overline{\lambda}_i} \tag{7.7}$$

2. We note in the above equation, the number of packets does not appear. This is important to note as in general packets encoded at the physical layer are sent as one block. If the block is erroneous, all packets within the subframe are erroneous.

7.2.2.2 Packet Delay

The packet delay can be found from Little's Theorem. Here, the average queueing delay can be given as

$$\mathcal{D}_i = \frac{\overline{q}_i}{\lambda_{q,i} T_f} \tag{7.8}$$

where $\overline{q}_i$ is the average queue size and $\lambda_{q,i}$ is the average enqueued arrival rate[3] for queue i. By design we can express $\overline{q}_i$ using the steady-state distribution $\theta_i(c_i, u_i|\mathbf{\Omega})$ as:

$$\overline{q}_i = \sum_{u_i \in \mathcal{U}_i} u_i \sum_{c_i \in \mathcal{C}_i} \theta_i(c_i, u_i|\mathbf{\Omega}) \tag{7.9}$$

and since $\lambda_{q,i}$ is also equal to the average service rate in steady-state, it can be expressed as

$$\lambda_{q,i} = \sum_{u_i \in \mathcal{U}_i} \sum_{c_i \in \mathcal{C}_i} \min(c_i, u_i) \theta_i(c_i, u_i|\mathbf{\Omega}) \tag{7.10}$$

We also note that $\lambda_{q,i}$ is related to the throughput as

$$\mathbb{E}_m[T_i(m)] = \lambda_{q,i} L_i \tag{7.11}$$

7.2.2.3 Per-Queue Transition Probability

The transition of each queue is solely based on the arrivals and departures from that queue. For poisson arrivals with average rate $\overline{\lambda}_i$ in packets per subframe, the transition probability is given as

$$p^{c_i}_{u_i;u_i'} = \begin{cases} Pr\left[A_i(m) = u_i' - [u_i - \min(u_i, c_i)]\right], u_i' < B_i \\ \sum_{j=B_i-[u_i-\min(u_i,c_i)]}^{\infty} Pr[A_i(m) = j], \qquad u_i' = B_i \end{cases} \tag{7.12}$$

3. Average rate at which packets enter the queue without being dropped.

where

$$Pr[A_i(m) = k] = \begin{cases} \frac{\overline{\lambda_i}^k \exp(-\overline{\lambda_i})}{k!}, & \text{if } k \geq 0 \\ 0, & \text{otherwise} \end{cases} \tag{7.13}$$

for a given average arrival rate $\overline{\lambda_i}$, buffer size B_i and all eligible packet service rates c_i. The transmission cost is found for each $c \in \mathcal{C}$. Let $\mathcal{P}(c)$ be the joint cost of choosing action c (action $c_1, c_2, \ldots c_K$ for each UE). As in Chapter 3, the marginal cost for each queue can be obtained using $\mathcal{P}(c)$ combined with the steady state distribution as in equation 3.24 of Chapter 3 to obtain the marginal cost function used in the MAC constrained optimization.

$\mathcal{P}(c)$ is dependant on the channel. Moreover, it is difficult to obtain a closed form expression on $\mathcal{P}(c)$ for all c in the SC-FDMA system due to the large number of resources. A near optimal uplink allocation algorithm is shown in Algorithm 2 where the closeness of the gap is given in Appendix D. In our work, we measure $\mathcal{P}(c)$ for all c by iteratively using the algorithm to obtain an approximation to $\mathcal{P}(c)$ for all $c \in \mathcal{C}$.

By design, the steady-state distribution $\theta_i(c_i, u_i|\mathbf{\Omega})$ must also satisfy the following balance property

$$\sum_{u_i' \in \mathcal{U}_i} \sum_{c_i' \in \mathcal{C}_i} \theta(c_i', u_i'|\mathbf{\Omega}) p^{c_i'}_{u_i', u_i} = \sum_{c_i \in \mathcal{C}_i} \theta(c_i, u_i|\mathbf{\Omega}), \forall u_i \tag{7.14}$$

7.2.3 Per Queue Objective Function

The computed power allocation found in the first component is used to derive the objective function for the local MAC layer rate assignment. First, the average marginal cost for taking an action c_i in queue 1 can be given as

$$\Upsilon_{1,x} = \sum_{c_2 \in \mathcal{C}_2} \ldots \sum_{c_K \in \mathcal{C}_K} P(x, c_2, \ldots, c_K) \cdot \pi_2(c_2|\mathbf{\Omega}) \times \ldots \times \pi_K(c_K|\mathbf{\Omega}) \tag{7.15}$$

Algorithm 2 Iterative Power Efficient Resource Allocation

$\mathcal{N} = \{1, 2, \ldots, M\}$
$\mathcal{N}_i = \emptyset, \quad \forall i \in \mathcal{K}$
$\mathcal{K}^{(a)} = \mathcal{K}$
$\mathcal{N}_i^{(f)} = \mathcal{N}, \quad \forall i \in \mathcal{K}$
while $|\mathcal{N}| > |\mathcal{K}^{(a)}|$ **do**
  for $i \in \mathcal{K}$ **do**
    if $\mathcal{N}_i \neq \emptyset$ **then**
      for $j \in \mathcal{N}_i^{(f)} \cap \mathcal{N}$ **do**
        $\Delta p_{i,j} = \alpha_i (P(\mathcal{N}_i, T_i, \boldsymbol{\gamma}) - P(\mathcal{N}_i \cup j, T_i, \boldsymbol{\gamma}))$
      end for
    else
      for $\mathcal{N}$ **do**
        $p_{i,j} = \alpha_i P(\mathcal{N}_i, T_i, \boldsymbol{\gamma})$
      end for
      $\Delta p_{i,j} = \min(\{p_{i,j}, j \in \mathcal{N} \setminus \arg\min_{j^* \in \mathcal{N}}(p_{i,j^*})\}) - \min(\{p_{i,j}, \forall j \in \mathcal{N}\})$
    end if
  end for
  if $\max(\Delta p_{i,j}) < 0$ **then**
    break
  end if
  $(i^*, j^*) = \arg\max_{i,j} \Delta p_{i,j}$
  $\mathcal{K}^{(a)} = \mathcal{K}^{(a)} \setminus i^*$
  $\mathcal{N}_{i^*} = \mathcal{N}_{i^*} \cup j^*$
  $\mathcal{N}_{i^*}^{(f)} = \{\min(\mathcal{N}_{i^*}) - 1, \max(\mathcal{N}_{i^*}) + 1\} \cap \mathcal{N}$
  $\mathcal{N} = \mathcal{N} \setminus j^*$
end while
if $|\mathcal{K}^{(a)}| \neq \emptyset$ **then**
  for $i \in \mathcal{K}^{(a)}$ **do**
    for $j \in \mathcal{N}$ **do**
      $p_{i,j} = \alpha_i P(\mathcal{N}_i, T_i, \boldsymbol{\gamma})$
    end for
  end for
  while $|\mathcal{K}^{(a)}| \neq \emptyset$ **do**
    $(i^*, j^*) = \arg\min_{i \in \mathcal{K}^{(a)}, j \in \mathcal{N}} p_{i,j}$
    $\mathcal{K}^{(a)} = \mathcal{K}^{(a)} \setminus i^*$
    $\mathcal{N}_{i^*} = j^*$
    $\mathcal{N} = \mathcal{N} \setminus j^*$
  end while
end if

where there are $i-1$ summations. Similar expressions can be found for all actions $c_i \in \mathcal{C}_i$ and found for all queues $k = 1, \ldots, K$ and where

$$\pi_i(x|\mathbf{\Omega}) = \sum_{u_i \in \mathcal{U}_i} \theta(x, u_i|\mathbf{\Omega}), x \in \mathcal{C}_i \tag{7.16}$$

$P(c_1, c_2, \ldots, c_K)$ denotes the average power allocated to transmit $\{c_1, c_2, \ldots, c_K\}$ packets for each queue i. In compact notation we denote this $\mathcal{P}(c)$ where each $c \in \mathcal{C}$ corresponds to a set $\{c_1, c_2, \ldots, c_K\}$ for all users.

7.2.4 Determining Power Applied Per State

Determination of $\mathcal{P}(c)$ is as follows. The bitrate T_i is easily obtained as $T_i = c_i L_i + L_{hdr}$ for that state c. Using $T_i, \forall i$ in Algorithm 2, one can obtain $\mathcal{N}_i$ for each iteration and find the weighted transmission power using (7.5). The value of $\mathcal{P}(c)$ is obtained by averaging the weighted transmission power over all realizations of the channel matrix. Since we assume the channel is continuous, and the number of resource blocks is large, it is difficult to do this for all realizations. For the rest of this work, we obtain $\mathcal{P}(c)$ by averaging over a finite number of random realizations for each c.

Consider the following example where there are 5 UEs, and 24 RBs and each user is transmitting 400 bits (for this state c, *i.e.,* $T_1 = T_2 = \ldots = T_5$) over one subframe. In Figure 7.2, we see a histogram of the actual applied power per iteration (instantaneous $\widetilde{\mathcal{P}(c)}$ over 10000 iterations) for this state c (where RBs and power are allocated at each iteration based on the channel CSI). For each c, $\mathcal{P}(c)$ is the mean value (*i.e.,* 0.1879 or $-7.26dB$ in this example). Similar values can be obtained for all other $c \in \mathcal{C}$.

In general these values can be calculated in advance and stored. The proposed method is functional for a small to moderate number of instantaneous users as the size of $\mathcal{C}$ is given as

$$|\mathcal{C}| = \prod_{i=1}^{K} (Z_i + 1) \tag{7.17}$$

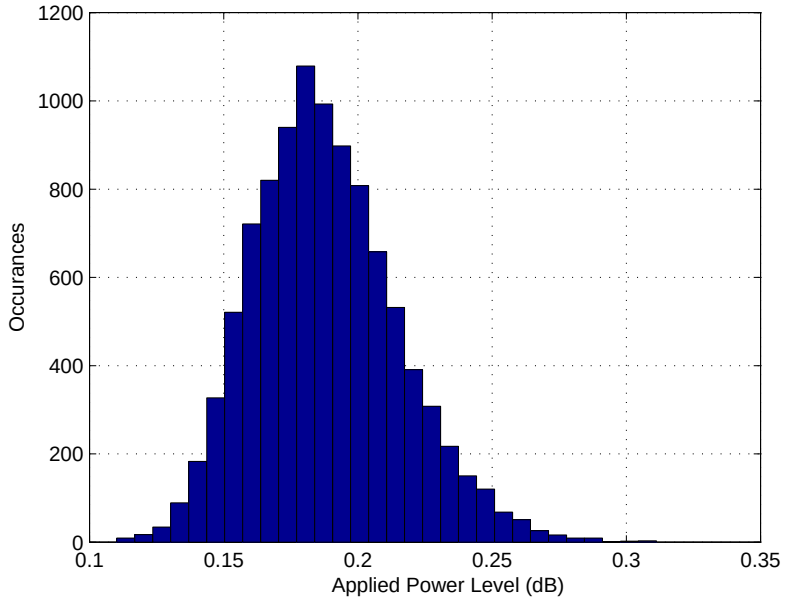


Figure 7.2: Histogram of Power Applied per Iteration

For the case above, where we assume $Z_i = 4$, we see that $|\mathcal{C}| = 5^{(4+1)} = 3125$. While we observe that the number of states increases exponentially, it is still considerably more efficient than consideration of the joint buffer occupancy states. For example, if each user had a buffer size of up to 25 SDUs, the resulting space would be $5^{(25+1)} = 1.49 \times 10^{18}$.

7.2.5 Iterative Policy Solver

The above steady-state action probabilities are coupled through the policy $\mathbf{\Omega}$. The value $P(c)$ is the total power associated with taking actions c_1 through c_K in each queue (or one for each state $c \in \mathcal{C}$) found earlier. Here we need to highlight that the above expression contains the steady-state probability of choosing an action in each queue. The result of which implies that it is not possible to directly decouple and consider each queue independently. As a result, we employ the per-queue iterative policy solver developed in Chapter 3. The solver operates as follows.

Firstly, each queue policy vector is initialized to

$$\pi_i(x|\mathbf{\Omega}^{(0)}) = \frac{1}{Z_i + 1}, \forall x, i \tag{7.18}$$

where $\mathbf{\Omega}^{(n)}$ denotes the policy computed at iteration n. Let $\mathcal{K}$ denote the set of users where $\mathcal{K} = \{1, 2, \ldots, K\}$ and n denote the iteration number where initially $n = 1$. At each iteration, $i^* = (n \mod K) + 1$ where n is incremented for each iteration. For each iteration we solve for $\pi_{i^*}(x|\mathbf{\Omega}^{(n)})$ as follows.

As in Chapter 3, the constrained MDP problem at each iteration is solved using Linear Programming (LP) techniques formulated as $\arg\min_{\mathbf{x}} \mathbf{c}^T\mathbf{x}$, subject to $\mathbf{A}\mathbf{x} \leq \mathbf{b}$, $\mathbf{A}_{eq}\mathbf{x} = \mathbf{b}_{eq}$, $\mathbf{x} \geq 0$ where $\mathbf{A}$ and $\mathbf{A}_{eq}$ are matrices and $\mathbf{x}, \mathbf{b}, \mathbf{b}_{eq}$ and $\mathbf{c}$ are column vectors. The vector $\mathbf{x}$ is the solution to the optimization problem. In our problem, the elements are given as

$$\mathbf{x} = [\boldsymbol{\theta}_{i^*}(\mathcal{C}_{i^*}, 0|\mathbf{\Omega}^{(n)}), \ldots, \boldsymbol{\theta}_{i^*}(\mathcal{C}_{i^*}, B_{i^*}|\mathbf{\Omega}^{(n)})]^T \tag{7.19}$$

with each $\boldsymbol{\theta}_{i^*}(\mathcal{C}_{i^*}, u_{i^*}|\mathbf{\Omega}^{(n)})$ being a row vector with entries for each $c_{i^*} = 0, 1, \ldots, Z_{i^*}$.

The objective function is of the form $\mathbf{c}^T\mathbf{x}$. The vector $\mathbf{c}$ is comprised of the total power cost for taking an action. Each entry of $\mathbf{c}$ corresponds to the entry in $\mathbf{x}$ with the value of entries in $\mathbf{c}$ given by $\Upsilon_{i^*,c_{i^*}}$ in (7.15).

$$\mathbf{c} = [\underbrace{\Upsilon_{i^*,1}, \ldots, \Upsilon_{i^*,Z_{i^*}+1}}_{1}, \overbrace{\ldots}^{2 \ldots B_{i^*}}, \underbrace{\Upsilon_{i^*,1}, \ldots, \Upsilon_{i^*,Z_{i^*}+1}}_{B_{i^*}+1}] \tag{7.20}$$

The equality constraints are comprised of the balance equations and the causality constraint (total probability space). In matrix form, the balance equations can be

expressed as $\mathbf{P} \times \mathbf{x} = \mathbf{\Phi}_0 \times \mathbf{x}$ where $\mathbf{P}$ is given by

$$\mathbf{P} = \begin{bmatrix} \mathbf{p}_{0;0}^{\mathcal{C}_{i^*}} & \cdots & \cdots & \mathbf{p}_{B_{i^*};0}^{\mathcal{C}_{i^*}} \\ \vdots & \mathbf{p}_{1;1}^{\mathcal{C}_{i^*}} & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{p}_{0;B_{i^*}}^{\mathcal{C}_{i^*}} & \cdots & \cdots & \mathbf{p}_{B_{i^*};B_{i^*}}^{\mathcal{C}_{i^*}} \end{bmatrix} \tag{7.21}$$

with $\mathbf{p}_{q;q'}^{\mathcal{C}_{i^*}}$ as a $1 \times (Z_{i^*} + 1)$ row vector with entries

$$\mathbf{p}_{q;q'}^{\mathcal{C}_{i^*}} = [p_{q;q'}^{1}, \ldots, p_{q;q'}^{Z_{i^*}+1}] \tag{7.22}$$

and the quantity $\mathbf{\Phi}_0$ is given as the $B_{i^*} + 1$ row matrix

$$\mathbf{\Phi}_0 = \begin{bmatrix} \mathbf{1}_{1\times(Z_{i^*}+1)} & 0 & \cdots & 0 \\ 0 & \mathbf{1}_{1\times(Z_{i^*}+1)} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{1}_{1\times(Z_{i^*}+1)} \end{bmatrix} \tag{7.23}$$

Combining the above with the causality constraint on the total probability space we have our overall equality constraints given as

$$\mathbf{A}_{eq} = \begin{bmatrix} \mathbf{P} - \mathbf{\Phi}_0 \\ \mathbf{1}_{1\times((Z_{i^*}+1)(B_{i^*}+1))} \end{bmatrix} \tag{7.24}$$

$$\mathbf{b}_{eq} = [\mathbf{0}_{1\times(B_{i^*}+1)} \quad 1]^T \tag{7.25}$$

The inequality constraints are used to describe the throughput and delay con-

---

$$\mathbf{U} = [\min(0,0), \min(1,0), \ldots, \min(Z_{i^*},0), \min(0,1), \ldots, \min(Z_{i^*},B_{i^*})] = -\mathbf{w}_1 \tag{7.30}$$

---

straints. These constraints are given in two parts as

$$\mathbf{A} = \begin{bmatrix} \mathbf{w}_1 \\ \mathbf{w}_2 \end{bmatrix} \qquad \mathbf{b} = \begin{bmatrix} \mathbf{z}_1 \\ \mathbf{z}_2 \end{bmatrix} \tag{7.26}$$

where $\mathbf{w}_1$ is given as

$$\mathbf{w}_1 = -[\boldsymbol{\chi}_{i^*:n}(\mathcal{C}_{i^*}, 0), \ldots, \boldsymbol{\chi}_{i^*:n}(\mathcal{C}_{i^*}, B_{i^*})] \tag{7.27}$$

where $\boldsymbol{\chi}_{i^*:n}(\mathcal{C}_{i^*}, u_{i^*})$ is a row vector with entries $\chi_{i^*:n}(c_{i^*}, u_{i^*})$ for all $c_{i^*} \in \mathcal{C}_{i^*}$ and $\mathbf{z}_1$ is given as

$$\mathbf{z}_1 = -\bar{\lambda}_{i^*}(1 - P_{drop,i^*})T_f \tag{7.28}$$

Finally, $\mathbf{w}_2$ is given as

$$\mathbf{w}_2 = \mathbf{Q} \times \boldsymbol{\Phi}_0 - \mathcal{D}_{i^*}\mathbf{U} \qquad \mathbf{z}_2 = 0 \tag{7.29}$$

where $\mathbf{Q} = [0, 1, \ldots, B_{i^*}]$ and $\mathbf{U}$ is given in (7.30).

The total system power at iteration n is given as

$$\mathcal{P}^{(n)} = \sum_{c_1 \in \mathcal{C}_1} \cdots \sum_{c_K \in \mathcal{C}_K} P(c_1, c_2, \ldots, c_K) \cdot \pi_1(c_1|\boldsymbol{\Omega}^{(n)}) \times \ldots \times \pi_K(c_K|\boldsymbol{\Omega}^{(n)}) \tag{7.31}$$

The process continues iteratively until reaching one of the following stopping conditions

$$(i) \quad |\mathcal{P}^{(n-k)} - \mathcal{P}^{(n-k-1)}| < \epsilon, k = 0, \ldots, K-1, n \geq K$$
$$(ii) \quad n > MAX_{iter}$$

where MAX_{iter} is the preset maximum number of iterations and ϵ is a small positive number. The above convergence details were presented in Chapter 3.

7.3 Simulation Results and Discussion

Results are presented with universal parameters summarized in Table 7.1 with the system modelled in MATLAB. Details are provided for the least-squares approximation as well as the applied power versus various parameters.

7.3.1 Least-Squares Applied Power Approximation

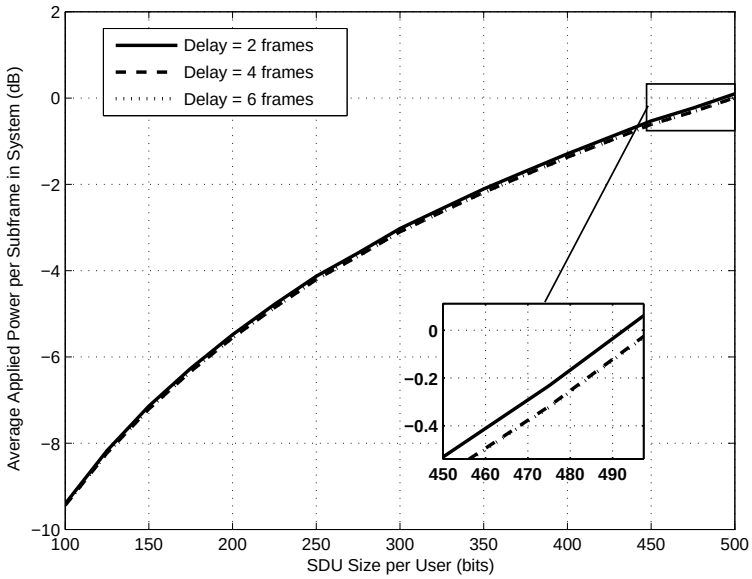
The justification behind use of the pre-described fit function was shown in Figure 6.3 in Chapter 6. Here we see in this comparison for $2, 4$, and 8 RBs with $BLER_{tgt} = 10\%$. The least-squares approximation is shown to hold tightly to the actual BLER function; providing for a more tractable computation of the required SNR level and justifying its use as a suitable alternative in computation of the required SNR.

7.3.2 Average Applied Power

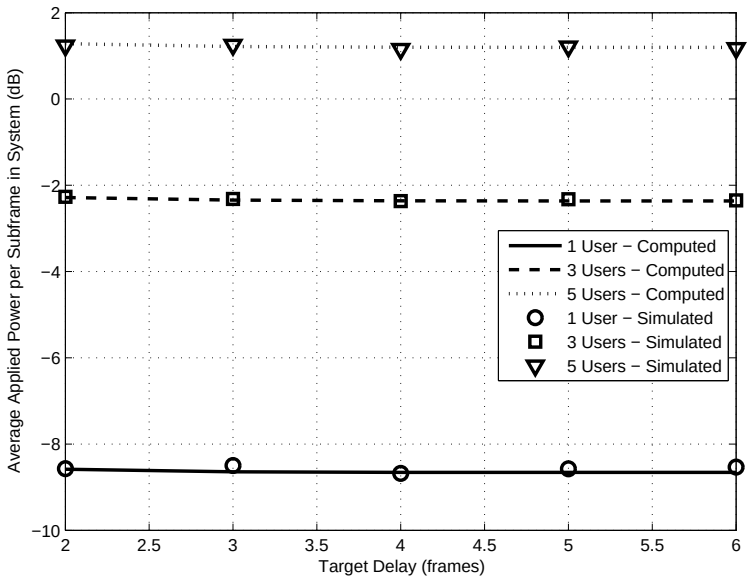
Results for the average applied power is shown in Figure 7.3. From these we observe several trends. Firstly, we observe a very low impact on the delay beyond 2 frames in terms of percentage difference. This finding is consistent with the results found in previous chapters. The impact on SDU size is significantly larger in terms of average power applied. We also observe the impact on the number of users in Figure 7.3(b).

Table 7.1: Simulation Parameters

Parameter	Value	Parameter	Value
γ_0	10dB	L_{hdr}	30 bits
N_{sym}	14	α_i	$1 \quad \forall i$
N_{ctrl}	3	$P_{drop,i}$	$0.1\% \quad \forall i$
$BLER_{tgt}$	10%	B_i	25 packets $\quad \forall i$
Number of Subframes	10000	$\bar{\lambda}_i$	1.5 packets per frame, $\quad \forall i$
T_f	1ms	D_i	4 frames $\quad \forall i$
Number of RBs (M)	24	Z_i	$4 \quad \forall i$
Number of UEs (K)	2	T_i	$150 + 50i, \quad \forall i$
ϵ	10^{-7}	MAX_{iter}	$10K$



(a) versus target average delay and SDU size



(b) versus target average delay and number of UEs

Figure 7.3: Average applied power per subframe versus system parameters

Here as expected the system expends additional power to accommodate the increase in users. In this graph the computed and simulated results are shown (computed

and given as the solution to the optimal policy, simulated obtained through implementation of the computed policy). The simulated results closely follow the expected scheduling policy performance.

7.4 Chapter Summary

In this chapter we designed and evaluated the design of an energy efficient scheduler for multiuser SC-FDMA uplink. By exploiting part of our previously designed iterative scheduling technique, with a near optimal iterative resource allocation mechanism, a low-complexity scheduling policy was obtained. The proposed design was compared versus various parameters.

Future work in this area can take a number of directions. For example, performance of the system described in this chapter depends highly on the allocation method at the PHY layer and as such, there is a need to develop low-complexity optimal or near optimal algorithms for resource block assignment for margin adaptive SC-FDMA. In the next chapter, details of this, and other related open problems are discussed under future work.

Chapter 8
Thesis Summary and Future Work

As time goes on, the need for energy efficient scheduling techniques becomes more and more paramount. Interestingly enough, while the penetration rate for mobile users and breadth of device features continues to increase, the need for more energy efficient communication is still of importance as still a large percentage of mobile device energy consumption is in the mobile radio.

8.1 Thesis Summary

In this thesis designs are proposed for both scheduling and resource allocation methods that offer reduced complexity in multi-stream (multi-user), multi-channel (MS-MC) scenarios in both downlink and uplink systems. In Chapter 3 we examined how we can apply MDP theory to a multi-queue environment using MIMO SVD in a way that reduces complexity and exploits QSI to minimize transmission energy. This novel approach can be used to efficiently schedule traffic streams with differing QoS parameters without needing to result to full scale optimization techniques while still guaranteeing throughput and average delay constraints. While we present some novel approaches, the work in this chapter was limited to the case of static channels.

However, next in Chapters 4 and 5 this static assumption was relaxed. First, the main contributions in Chapter 4 resulted in a detailed study of sparse MIMO channels. This research work provided a better understanding of the time-average and time-varying statistics in this class of channels. This information was used in Chapter 5 to design an extended scheduling scheme that exploited this information about the underlying channel statistics. The main contribution of this Chapter 5 was

an extension of the work in Chapter 3 to utilize coded modulation schemes and relaxation of the static channel assumption. Further, the important tradeoff of complexity and energy efficiency in reference to channel partitioning was also examined.

In Chapters 6 and 7, we turned to the multi-user uplink scenario. This work focused on the heterogeneous multiuser case as the incoming traffic streams and SC-FDMA as the multiple channel medium. Unlike the MIMO SVD system, SC-FDMA introduced additional complexity considerations. In these chapters, the complexity considerations were able to be exploited in a way that effectively reduced the computational complexity. Specifically in Chapter 6 we proposed a way to reduce the scheduling intervals when synchronous HARQ was considered by exploiting its periodicity. We further proposed reduced complexity methods and compared these with the optimal per ARQ slot allocation. Chapter 7 relaxed the HARQ constraints and looked at combining the novel work in Chapters 3 and 5 with the allocation mechanisms found in Chapter 6. In this way, the scheduler was able to exploit QSI to energy efficiently allocate users each subframe with only a priori knowledge of the channel distribution. Such knowledge could easily be obtained through measurement campaigns.

Overall, one of the primary arguments made in this thesis is the large complexity associated with heterogenous scheduling techniques, particularly when attempting to minimize energy expended in the mobile radio. While in this thesis, methods of complexity reduction in scheduling were shown while attempting to minimize energy expenditure, it is important to note that this research area is still wide open and that there is no single solution to this complexity problem. This is particularly true as the constraints imposed by practical systems vary dramatically from system to system.

8.2 Future Work

The work in this thesis presents some investigation into the multi-stream, multi-channel scheduling problem, particularly with a focus on complexity issues. While we have proposed some novel techniques to reduce complexity and address this problem,

there still remains numerous open issues in this area. In this section we highlight some of these major issues.

8.2.1 Outage and Partial Outage Events in MAC Rate Allocation

In Chapter 3, we described the problem associated with no limitation on the maximum instantaneous power. This problem was further complicated in Chapter 5. The need to characterize outage events (either due to low subchannel SNRs or exceeding the instantaneous sum-rate from all queues) is an important next step in extending our scheduling algorithm derived in these chapters.

One method of accomplishing this goal is considering an abstracted channel quality indicator in the MAC rate allocation stage while leaving the PHY allocation (channel, rate and power) to operate independently as before. This could be done as follows. Let $\theta_i(u_i, c_i, \varrho|\mathbf{\Omega})$ be the steady-state probability of queue i existing in a state where u_i is the number of packets in the queue, c_i is the action taken and ϱ is the state of the physical layer for a given policy $\mathbf{\Omega}$.

The state of the physical can be described in any way. In fact, since the PHY layer allocation is handled to minimize average allocated transmission power locally, this additional coupling parameter can be used to describe the PHY layer in detail or almost completely abstracted. In the case of handling the outage problem discussed previously, the parameter can take on the values $\varrho \in \mathcal{V} = \{0, 1\}$ where 0 denotes an outage event, and 1 denotes a non-outage event. In this case, it only doubles the state-space of the locally optimized queue rate allocation. Transition probabilities for ϱ are independent of the MAC policy $\mathbf{\Omega}$.

It is also possible to have a multi-state $\varrho \in \mathcal{V}$ which can provide less abstraction of the physical layer. Unlike CSI or partial CSI, ϱ can be used to quantify any part of the physical layer at the MAC. The benefit of this approach is that it provides granularity control over complexity. Research into $\mathcal{V}$ beyond outage information can be the study of future research projects.

8.2.2 Availability and Quantity of Channels

In all chapters throughout this thesis, one of the assumptions is the known quantity of available channels for use (either through MIMO or resources in frequency with SC-FDMA). It may arise however that the number of channel resources may be random and change in time. Such problems can arise in a number of ways, such as with MIMO as scattering conditions change dramatically in time. It is therefore important to extend some of the proposed methods in this thesis to the case where the number of channels are random in future work.

8.2.3 Ordered Statistics for Sparse MIMO

In Chapter 4, we derived statistics for the unordered eigenvalues and SNR as well as the unordered level-crossing rates and average fade durations. While these are important measurements, it is also important to understand these measurements for ordered statistics, particularly for use in resource allocation.

Recently, a method was proposed in [84] to determine these level crossing statistics for ordered random processes. It is possible that these results could be applied to the sparse MIMO scenario presented in Chapter 4 which would strengthen the understanding of their performance in terms of ordered statistics and time-varying parameters, as well as be beneficial in future scheduler designs.

8.2.4 Accurate Channel Modelling for BTFPDS

In Chapter 6, we assumed that the channel was static for the duration of an entire ARQ slot (block fading over an ARQ slot). While this may be true in the slow fading environment, this assumption may not hold in fast fading conditions. One method that could be used to combat this assumption is to employ some of the concepts of the BTFDPS allocation method in terms of admission control for users limited to one per ARQ slot. Further, the scheduler could employ estimates of the future channel of those users throughout the ARQ slot based on past and current channel measurements. Estimation accuracy in this case plays a vital role in the performance of the allocation as accuracy will decrease as a function of the subframe in the ARQ

slot and this accuracy, and its impact on performance should be studied in detail in future work.

8.2.5 Low Complexity Margin Adaptive Allocation for SC-FDMA

In Chapters 6 and 7 we presented a low complexity algorithm for margin adaptive (MA) resource allocation for localize SC-FDMA. The algorithm iteratively allocated frequency resource to maximize the *power level gain* per iteration. While this method was shown to perform close to optimal power allocation for smaller transport block (TB) sizes, this gap deviates for very large blocks.

While efficient MA methods have been studied heavily for OFDM, SC-FDMA has received significantly less attention. This is particularly important, as this technology has become popular in recent years due to its low peak-to-average power ratio (PAPR) as a solution for mobile devices. More importantly, it is these devices where MA is of greater importance due to their battery limited nature. It is therefore important to design low-complexity, near optimal MA algorithms that work under a wide range of conditions (*i.e.,* channels, TB sizes, etc) in future work.

8.2.6 Frequency Correlated Performance of SC-FDMA Allocation Methods

In Chapters 6 and 7, we assumed a block fading model where each channels were also independent as a function of user and RBs in frequency. In realistic conditions, there may be some correlation in frequency between the SNR of individual channels. Intuitively, such channels should improve the performance of our iterative allocation schemes as for a given initial allocation, there is a higher probability of finding a good quality adjacent channel. This however, may not be true as the same can be said for an initially poor selected channel. It is therefore important to study this impact on MA iterative allocation algorithms, particularly for localized SC-FDMA in future work.

References

[1] Q. Spencer, A. Swindlehurst, and M. Haardt, "Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels," *IEEE Transactions on Signal Processing*, vol. 52, no. 2, pp. 461–471, February 2004.

[2] W. Rhee and J. Cioffi, "Increase in capacity of multiuser OFDM system using dynamic subchannel allocation," in *Proc. of IEEE Vehicular Technology Conference, Spring. Tokyo Japan*, May 2000.

[3] G. Caire and S. Shamai, "On the achievable throughput of a multiantenna Gaussian broadcast channel," *IEEE Transactions on Information Theory*, vol. 49, no. 7, pp. 1691–1706, July 2003.

[4] M. Rashid, E. Hossain, and V. Bhargava, "Cross-layer analysis of downlink V-BLAST MIMO transmission exploiting multiuser diversity," *IEEE Transactions on Wireless Communications*, vol. 8, no. 9, pp. 4568–4579, September 2009.

[5] D. Niyato, E. Hossain, and D. Kim, "Joint admission control and antenna assignment for multiclass QoS in spatial multiplexing MIMO wireless networks," *IEEE Transactions on Wireless Communications*, vol. 8, no. 9, pp. 4855–4865, September 2009.

[6] D. J. Dechene and A. Shami, "Energy efficient quality of service traffic scheduler for MIMO downlink SVD channels," *IEEE Transactions on Wireless Communications*, vol. 9, no. 12, pp. 3750–3761, December 2010.

[7] D. J. Dechene, S. L. Primak, and A. Shami, "On the first and second order statistics of sparse MIMO channels," in *Proc. of 25th Queens Biennial Symposium on Communications. Kingston, Canada*, May 2010.

[8] D. J. Dechene and A. Shami, "QoS, channel and energy-aware packet scheduling over multiple channels," *IEEE Transactions on Wireless Communications*, vol. 10, no. 4, pp. 1058–1062, April 2011.

[9] D. J. Dechene and A. Shami, "Energy efficient resource allocation in SC-FDMA uplink with synchronous HARQ constraints," in *Proc. of IEEE International Conference on Communications. Kyoto, Japan*, June 2011.

[10] D. Buckingham, "Information-outage analysis of finite-length codes," Ph.D. dissertation, West Virginia University. http://wvuscholar.wvu.edu:8881/exlibris/dtl/d3_1/apache_media/13988.pdf, 2008.

[11] *IEEE Standard 802.11e-2005: Wireless LAN MAC and PHY Specifications. Amendment 8: MAC QoS Enhancements*, 2005.

[12] *International Standard [for] Local and Metropolitan Area Networks - Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems*, IEEE 802.16 WG, IEEE STD 802.16e-2005, 2005.

[13] Q. Liu, X. Wang, and G. Giannakis, "A cross-layer scheduling algorithm with QoS support in wireless networks," *IEEE Transactions on Vehicular Technology*, vol. 55, no. 3, p. 839, May 2006.

[14] K. Wongthavarawat and A. Ganz, "Packet scheduling for QoS support in IEEE 802.16 broadband wireless access systems," *International Journal of Communication Systems*, vol. 16, no. 1, pp. 81–96, February 2003.

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

[16] P. Wolniansky, G. Foschini, G. Golden, and R. Valenzuela, "V-BLAST: an architecture for realizing very high data rates overthe rich-scattering wireless channel," *URSI International Symposium on Signals, Systems, and Electronics. Pisa, Italy*, pp. 295–300, September 1998.

[17] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge University Press, 2005.

[18] V. Tsibonis, L. Georgiadis, and L. Tassiulas, "Exploiting wireless channel state information for throughput maximization," *IEEE Transactions on Information Theory*, vol. 50, no. 11, pp. 2566–2582, November 2004.

[19] Q. Zhang and S. Kassam, "Finite-state Markov model for Rayleigh fading channels," *IEEE Transactions on Communications*, vol. 47, no. 11, pp. 1688–1692, November 1999.

[20] Q. Liu, S. Zhou, and G. Giannakis, "Queuing with adaptive modulation and coding over wireless links: cross-layer analysis and design," *IEEE Transactions on Wireless Communications*, vol. 4, no. 3, pp. 1142–1153, May 2005.

[21] X. Bai and A. Shami, "Two dimensional cross-layer optimization for packet transmission over fading channel," *IEEE Transactions on Wireless Communications*, vol. 7, no. 10, pp. 3813–3822, October 2008.

[22] R. Berry and R. Gallager, "Communication over fading channels with delay constraints," *IEEE Transactions on Information Theory*, vol. 48, no. 5, pp. 1135–1149, 2002.

[23] M. Neely, "Optimal Energy and Delay Tradeoffs for Multiuser Wireless Downlinks," *IEEE Transactions on Information Theory*, vol. 53, no. 9, pp. 3095–3113, September 2007.

[24] R. Bellman, *Dynamic programming*. Dover Publisher, 2003.

[25] *IEEE Standard 802.11-1999: Wireless LAN MAC and PHY Specifications*, 1999.

[26] Q. Zhang and S. Kassam, "Finite-state Markov model for Rayleigh fading channels," *IEEE Transactions on Communications*, vol. 47, no. 11, pp. 1688–1692, November 1999.

[27] J. Razavilar, K. Liu, and S. Marcus, "Jointly optimized bit-rate/delay control policy for wireless packet networks with fading channels," *IEEE Transactions on Communications*, vol. 50, no. 3, pp. 484–494, March 2002.

[28] D. Bertsekas, R. Gallager, and T. Nemetz, *Data networks*. Prentice-hall Englewood Cliffs, NJ, 1987.

[29] S.-H. Kuo and J. Cavers, "Energy optimal scheduler for diversity fading channels with maximum delay constraints," *IEEE Transactions on Wireless Communications*, vol. 8, no. 11, pp. 5520–5529, November 2009.

[30] *Information technology Generic Coding of Moving Pictures and Associated Audio Information: Systems*, ISO/IEC 13818-1:2000, 2000.

[31] T. F. Maciel and A. Klein, "On the performance, complexity, and fairness of suboptimal resource allocation for multiuser MIMO-OFDMA systems," *IEEE Transactions on Vehicular Technology*, vol. 59, no. 1, pp. 406–419, January 2010.

[32] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, achievable rates, and sum-rate capacity of Gaussian MIMO broadcast channels," *IEEE Transactions on Information Theory*, vol. 49, no. 10, pp. 2658–2668, October 2003.

[33] M. Costa, "Writing on dirty paper," *IEEE Transactions on Information Theory*, vol. 29, no. 3, pp. 439–441, March 1983.

[34] C. Windpassinger, R. Fischer, T. Vencel, and J. Huber, "Precoding in multiantenna and multiuser communications," *IEEE Transactions on Wireless Communications*, vol. 3, no. 4, pp. 1305–1316, April 2004.

[35] Z. Tu and R. Blum, "Multiuser diversity for a dirty paper approach," *IEEE Communications Letters*, vol. 7, no. 8, pp. 370–372, August 2003.

[36] H. Viswanathan, S. Venkatesan, and H. Huang, "Downlink capacity evaluation of cellular networks with known-interference cancellation," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 5, pp. 802–811, June 2003.

[37] G. Dimic and N. Sidiropoulos, "On downlink beamforming with greedy user selection: performance analysis and a simple new algorithm," *IEEE Transactions on Signal Processing*, vol. 53, no. 10 Part 1, pp. 3857–3868, October 2005.

[38] T. Yoo and A. Goldsmith, "Optimality of zero-forcing beamforming with multiuser diversity," in *Proc of IEEE International Conference on Communications. Seoul, Korea*, May 2005, pp. 542–546.

[39] Z. Shen, R. Chen, J. Andrews, R. Heath Jr, and B. Evans, "Low complexity user selection algorithms for multiuser MIMO systems with block diagonalization," *IEEE Transactions on Signal Processing*, vol. 54, no. 9, pp. 3658–3663, September 2006.

[40] N. Jindal, W. Rhee, S. Vishwanath, S. Jafar, and A. Goldsmith, "Sum power iterative water-filling for multi-antenna Gaussian broadcast channels," *IEEE Transactions on Information Theory*, vol. 51, no. 4, pp. 1570–1580, April 2005.

[41] M. Kobayashi and G. Caire, "An iterative water-filling algorithm for maximum weighted sum-rate of Gaussian MIMO-BC," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 8, pp. 1640–1646, August 2006.

[42] K. Seong, M. Mohseni, and J. Cioffi, "Optimal resource allocation for OFDMA downlink systems," in *Proc of IEEE International Symposium on Information Theory. Seattle, USA*, December 2006, pp. 1394–1398.

[43] L. Georgiadis, M. Neely, M. Neely, and L. Tassiulas, *Resource allocation and cross layer control in wireless networks*. Now Publishers Inc., 2006.

[44] C. Manikandan, S. Bhashyam, and R. Sundaresan, "Cross-layer scheduling with infrequent channel and queue measurements," *IEEE Transactions on Wireless Communications*, vol. 8, no. 12, pp. 5737–5742, December 2009.

[45] K. Kar, X. Luo, and S. Sarkar, "Throughput-optimal scheduling in multichannel access point networks under infrequent channel measurements," in *Proc of. IEEE International Conference on Computer Communications. Anchorage, USA*, May 2007, pp. 1640–1648.

[46] S. Kittipiyakul and T. Javidi, "Resource allocation in OFDMA with time-varying channel and bursty arrivals," *IEEE Communications Letters*, vol. 11, no. 9, pp. 708–710, September 2007.

[47] V. Lau and Y. Cui, "Delay-optimal power and subcarrier allocation for OFDMA systems via stochastic approximation," *IEEE Transactions on Wireless Communications*, vol. 9, no. 1, pp. 227–233, January 2010.

[48] V. Lau and Y. Chen, "Delay-optimal power and precoder adaptation for multistream MIMO systems," *IEEE Transactions on Wireless Communications*, vol. 8, no. 6, pp. 3104–3111, June 2009.

[49] A. Goldsmith and S. Chua, "Variable-rate variable-power MQAM for fading channels," *IEEE Transactions on Communications*, vol. 45, no. 10, pp. 1218–1230, October 1997.

[50] X. Zhu, J. Huo, X. Xu, C. Xu, and W. Ding, "QoS-Guaranteed scheduling and resource allocation algorithm for IEEE 802.16 OFDMA system," in *Proc of IEEE International Conference on Communications. Beijing, China*, May 2008, pp. 3463–3468.

[51] X. Bai, A. Shami, and S. Primak, "Optimal power control over fading channel with cross-layer performance constraint," in *Proc. of IEEE International Conference on Communications. Beijing, China*, May 2008, pp. 2265–2269.

[52] C. Wang, P.-C. Lin, and T. Lin, "A cross-layer adaptation scheme for improving IEEE 802.11e QoS by learning," *IEEE Transactions on Neural Networks*, vol. 17, no. 6, pp. 1661–1665, November 2006.

[53] F. Meshkati, H. Poor, S. Schwartz, and R. Balan, "Energy-efficient resource allocation in wireless networks with quality-of-service constraints," *IEEE Transactions on Communications*, vol. 57, no. 11, pp. 3406–3414, November 2009.

[54] R. Louie, M. Mckay, and I. Collings, "Maximum sum-rate of MIMO multiuser scheduling with linear receivers," *IEEE Transactions on Communications*, vol. 57, no. 11, pp. 3500–3510, November 2009.

[55] R. Bultitude, G. Brussaard, M. Herben, and T. Willink, "Radio channel modelling for terrestrial vehicular mobile applications," in *In Proc. Millenium Conference on Antennas and Propagation. Davos, Switzerland*, April 2000, pp. 1–5.

[56] A. Burr, "Capacity bounds and estimates for the finite scatterers MIMO wireless channel," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 5, pp. 812– 818, May 2003.

[57] S. Chung and A. Goldsmith, "Degrees of freedom in adaptive modulation: a unified view," *IEEE Transactions on Communications*, vol. 49, no. 9, pp. 1561–1571, September 2001.

[58] O. Brun and J. Garcia, "Analytical solution of finite capacity M/D/1 queues," *Journal of Applied Probability*, vol. 37, no. 4, pp. 1092–1098, December 2000.

[59] X. Bai, A. Shami, and S. Primak, "Power efficient scheduling over fading channel for cross-layer optimization," *In Press, Wiley Wireless Communications and Mobile Computing*, 2011.

[60] P.-H. Kuo and P. Smith, "Temporal behavior of MIMO channel quality metrics," in *Proc. of International Conference on Wireless Networks, Communications and Mobile Computing. Maui, USA*, June 2005.

[61] P. Smith, P.-H. Kuo, and L. Garth, "Level crossing rates for MIMO channel eigenvalues: implications for adaptive systems," in *Proc. of IEEE International Conference on Communications. Seoul, Korea*, May 2005, pp. 2442–2446.

[62] A. Giorgetti, P. Smith, M. Shafi, and M. Chiani, "MIMO capacity, level crossing rates and fades: the impact of spatial/temporal channel correlation," *Journal of communications and networks*, vol. 5, no. 2, pp. 104–115, April 2003.

[63] S.Vaihunthan, S.Haykin, and M.Sellathurai, "MIMO channel capacity modeling using Markov models," in *Proc of IEEE Vehicular Technology Conference - Spring. Stockholm, Sweden*, May 2005, pp. 126–130.

[64] A. Paulraj, R. Nabar, and D. Gore, *Introduction to Space-Time Wireless Communications*. Cambridge, UK: Cambridge University Press, 2003.

[65] S. Primak and E. Sejdic, "Application of multitaper analysis to wireless communications problems," in *Proc. of ISABEL. Aalborg, Denmark*, October 2008.

[66] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 3rd ed. Boston, MA: McGraw-Hill, 1991.

[67] A. Paulraj, R. Nabar, and D. Gore, *Introduction to Space-Time Wireless Communications*. Cambridge University Press, 2003.

[68] R. Bapat and M. Beg, "Order statistics for nonidentically distributed variables and permanents," *Sankhyā: The Indian Journal of Statistics, Series A*, pp. 79–93, 1989.

[69] B. Hochwald, T. Marzetta, and V. Tarokh, "Multiple-antenna channel hardening and its implications for rate feedback and scheduling," *IEEE Transactions on Information Theory*, vol. 50, no. 9, pp. 1893–1909, September 2004.

[70] M. Abramowitz and I. Stegun, Eds., *Handbook of Mathematical Functions*. New York: Dover, 1965.

[71] D. da Costa and S. Aïssa, "Capacity analysis of cooperative systems with relay selection in Nakagami-m fading," *IEEE Communication Letters*, vol. 13, no. 9, pp. 637–639, September 2009.

[72] A. Tulino and S. Verdú , *Random Matrix Theory and Wireless Communications*. Now Publishers Inc, 2004.

[73] P. Ivanis, D. Drajic, and B. Vucetic, "Level crossing rates of Ricean MIMO channel eigenvalues for imperfect and outdated CSI," *IEEE Communications Letters*, vol. 11, no. 10, pp. 775–777, October 2007.

[74] M. Luccini, A. Shami, and S. Primak, "Cross-layer Optimisation of Network Performance over Multiple-Input Multiple-Output Wireless Mobile Channels," *IET Communications,*, vol. 4, no. 6, pp. 683–696, April 2010.

[75] 3GPP, "Evolved Universal Terrestrial Radio Access (E-UTRA); Long Term Evolution (LTE) physical layer; General description," TS 36.201.

[76] 3GPP, "Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer procedures," TS 36.213.

[77] F. Calabrese, "Scheduling and link adaptation for uplink SC-FDMA systems," Ph.D. dissertation, Aalborg University. http://vbn.aau.dk/files/19156393/PhD_Thesis_FrancescoDavideCalabrese_final_print.pdf, 2009.

[78] J. Ikuno, M. Wrulich, and M. Rupp, "Performance and modeling of LTE H-ARQ," in *Proc. of IEEE Workshop on Smart Antennas. Berlin, Germany*, February 2009, pp. 1–6.

[79] F. Calabrese, P. Michaelsen, C. Rosa, M. Anas, C. Castellanos, D. Villa, K. Pedersen, and P. Mogensen, "Search-tree based uplink channel aware packet scheduling for UTRAN LTE," in *Proc. of IEEE Vehicular Technology Conference, Spring. Marina Bay, Singapore*, May 2008, pp. 1949–1953.

[80] F. Calabrese, C. Rosa, M. Anas, P. Michaelsen, K. Pedersen, and P. Mogensen, "Adaptive Transmission Bandwidth Based Packet Scheduling for LTE Uplink," in *Proc. of IEEE Vehicular Technology Conference, Fall. Calgary, Canada*, September 2008, pp. 1–5.

[81] S. Jungsup, G. Gye-Tae, and K. Dong-Hoi, "Packet-scheduling algorithm by the ratio of transmit power to the transmission bits in 3GPP LTE downlink," *EURASIP Journal on Wireless Communications and Networking*, vol. 2010.

[82] M. Al-Rawi, R. Jantti, J. Torsner, and M. Sagfors, "On the performance of Heuristic opportunistic scheduling in the uplink of 3G LTE networks," in *Proc.*

of IEEE Personal, Indoor and Mobile Radio Communications. Cannes, France, September 2008, pp. 1–6.

[83] I. Wong, O. Oteri, and W. Mccoy, "Optimal resource allocation in uplink SC-FDMA systems," *IEEE Transactions on Wireless Communications*, vol. 8, no. 5, pp. 2161 –2165, May 2009.

[84] P. Dharmawansa, M. McKay, and P. Smith, "Analysis of the level crossing rates for ordered random processes," in *Proc of IEEE International Conference on Communications. Kyoto, Japan*, June 2011, pp. 1–5.

[85] S. Rice, *Statistical properties of a sine wave plus random noise.* Bell Telephone Laboratories, 1948.

[86] H. Wang and N. Moayeri, "Finite-state Markov channel-a useful model for radio communication channels," *IEEE Transactions on Vehicular Technology*, vol. 44, no. 1, pp. 163–171, February 1995.

[87] The MathWorks Inc., "MATLAB R2008a," 2008.

[88] M. Kozlov, S. Tarasov, and L. Khachiyan, "Polynomial solvability of convex quadratic programming," in *Soviet Mathematics Doklady*, vol. 20, no. 5, 1979, pp. 1108–1111.

[89] E. Dijkstra, "A note on two problems in connexion with graphs," *Numerische mathematik*, vol. 1, no. 1, pp. 269–271, 1959.

[90] D. White, *Markov Decision Processes.* John Wiley & Sons, 1992.

Appendix A
Probability Theory and Finite State Markov Channels

Throughout this thesis, there are a number of mathematical tools employed. For clarity, this chapter overview various mathematical tools, fundamentals and definitions that are used throughout this thesis including probability theory, finite state Markov chains and optimization techniques/formulations.

A.1 Probability Theory and Distributions

Let X be a real-valued, non-negative, continuous random variable (RV). This implies that for any realization of this random variable, the value of X will fall in $(0, \infty)$.The probability density function (PDF) of this random variable X is denoted $f_X(x)$ and completely describes the random variable. For any non-negative random variable, the PDF must satisfy

$$\int_0^\infty f_X(x)dx = 1 \tag{A.1}$$

As $f_X(x)$ denotes the PDF, the probability of a continuous RV taking on a value between a and b (inclusive) is given as

$$\Pr[a \leq X \leq b] = \int_a^b f_X(x)dx \tag{A.2}$$

implying the probability of a continuous RV taking on a specific value is 0:

$$\Pr[X = b] = \Pr[b < X \leq b] = \int_b^b f_X(x)dx = 0 \tag{A.3}$$

The above only holds true for continuous random variables. For discrete, or discontinuous RVs, the density function is not continuous and therefore a RV may take on a specific value with a probability greater than 0.

In addition to the PDF, the cumulative distribution function (CDF) also describes the RV. The CDF by definition for a non-negative RV is given as

$$F_X(x) = \Pr[X \leq x] = \int_0^x f_X(y)dy \tag{A.4}$$

where $f_X(y)$ is the PDF derived above. It's easy to observe that $F_X(0) = 0$, $F_X(\infty) = 1$.

A.2 Moments

The m^{th} moment of an RV X is given as

$$\mathbb{E}_X[X^m] = \int_0^\infty x^m f_X(x)dx \tag{A.5}$$

where the definition of expectation function $\mathbb{E}[\cdot]$ with respect to random variable X is

$$\mathbb{E}_X[g(X)] = \int_0^\infty g(x) f_X(x)dx \tag{A.6}$$

The first moment of an RV is known as the mean which we denote $\bar{\mu}_X$. Similarly, the m^{th} central moment of a random variable is the m^{th} moment about the mean, namely

$$\mathbb{E}_X[(X - \bar{\mu}_X)^m] = \int_0^\infty (x - \bar{\mu}_X)^m f_X(x)dx \tag{A.7}$$

The variance, or the second central moment, is a commonly used central moment to describe an RV. Frequently it is denoted as σ_X^2. Alternatively, it is easy to show that

the variance is also

$$\sigma_X^2 = \mathbb{E}_X[X^2] - (\mathbb{E}_X[X])^2 \tag{A.8}$$

For a Gaussian random variable following the distribution

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma_X^2}} \exp\left(-\frac{(x-\bar{\mu}_X)^2}{2\sigma_X^2}\right) \tag{A.9}$$

only the mean and variance are required to completely describe the random variable. In later chapters, the notation $\mathcal{N}(\bar{\mu}_X, \sigma_X^2)$ is used to denote a Normal (or Gaussian) distribution with mean $\bar{\mu}_X$ and variance σ_X^2.

A.3 Second Order Statistics

A continuous time random process $X(t)$ is a process that evolves randomly in time. In general a random process is defined as [66] a family of time functions depending on a particular parameter(s). At any instant of time, *e.g.,* t_1, the process may be sampled. A sample of the random process at anytime such as t_1 is a random variable which we can denote as X_1.

In the interest of this thesis, Chapter 4 focuses on study of the evolution of sparse MIMO channels. Two well-known and important measures of a random process in the context of communication systems are its level crossing rate and average fade duration. These parameters are used frequency for scheduling techniques as well as determining burst errors in communication systems.

A.3.1 Level Crossing Rate (LCR)

The level crossing rate (LCR) of a process is a measure of the expected rate of a process crossing a threshold in a given direction (either positive or negative). In general, the LCR of a process provides an idea of how quickly a process evolves in time. The general formula for the LCR of a non-negative process over a threshold x

in one direction was originally found by Rice and is given as [85]

$$N_X(x) = \int_0^\infty \dot{x} f_{X,\dot{X}}(x, \dot{x}) d\dot{x} \tag{A.10}$$

where $f_{X,\dot{X}}(x, \dot{x})$ is the joint PDF of X and its time derivative process $\dot{X}$. In general, this joint PDF is relatively hard to obtain. In certain cases, and is often assumed, that the derivative process is independent of the process itself. As a result the joint PDF can be expressed as

$$f_{X,\dot{X}}(x, \dot{x}) = f_X(x) f_{\dot{X}}(\dot{x}) \tag{A.11}$$

From this, (A.10) becomes

$$N_X(x) = f_X(x) \int_0^\infty \dot{x} f_{\dot{X}}(\dot{x}) d\dot{x} \tag{A.12}$$

Finally, for a Gaussian process, $f_{\dot{X}}(\dot{x})$ is also a Gaussian RV with zero mean and variance $\sigma^2_{\dot{X}}$. In this case Rice showed that the LCR is given as

$$N_X(x) = f_X(x) \sqrt{\frac{\sigma^2_{\dot{X}}}{2\pi}} \tag{A.13}$$

where for a Gaussian process

$$\sigma^2_{\dot{X}} = - \frac{\partial^2}{\partial \tau^2} \{\rho_X(\tau)\} \Big|_{\tau=0} \tag{A.14}$$

and $\rho_X(\tau)$ is the time correlation function of process $X(t)$.

A.3.2 Average Fade Duration (AFD)

The average fade duration (AFD) of a process is a measure of the expected duration of time a process remains below a threshold during each instance that process is below

that threshold. The AFD effectively is

$$AFD = \frac{\text{Average Amount of Time Below a Threshold}}{\text{Number of Times Below A Threshold}} \tag{A.15}$$

which is easily given mathematically as

$$AFD(x) = \frac{\int_0^x f_X(y)dy}{N_X(x)} \tag{A.16}$$

A.4 Finite State Markov Channel

The idea of finite state Markov Channels (FSMC) was originally introduced in [86] and is known as the Wang-Moyaeri model. In general, channels measurements such as the SNR level of a wireless channel are continuous time quantities. Using FSMC, the channel can be considered as a finite number of states from an adaptation or resource allocation perspective. In this way, instead of performing adaptation on the continuous SNR of the channel, adaptation is solely based on the current state of the channel as well as information about transitions between states.

Let the instantaneous SNR of a channel be $\gamma(t)$ such that $0 < \gamma(t) < \infty, \forall t$ with SNR distribution $f_\gamma(\gamma)$. Let $\mathcal{S}$ be a set of K finite, non-overlapping states such that $\mathcal{S} = \{s_1, s_2, \ldots, s_K\}$. Each state is bounded by $[\varphi_k, \varphi_{k+1})$ with $\varphi_1 = 0$ and $\varphi_{K+1} = \infty$. For any time t, the $\gamma(t)$ falls within a single state $s_k \in \mathcal{S}$.

The quantities φ_k are known as the partition boundaries. The probability of the channel falling in state s_k is a function of these boundaries and given as

$$\Pr[s_k \in \mathcal{S}] = \int_{\varphi_k}^{\varphi_{k+1}} f_\gamma(\gamma)d\gamma \tag{A.17}$$

The transition probabilities between states can be computed in a manner such as that in [19]. We denote the frame of interest having a duration T_f. If the changes in-time are assumed to be small compared to the frame duration of interest (T_f) then

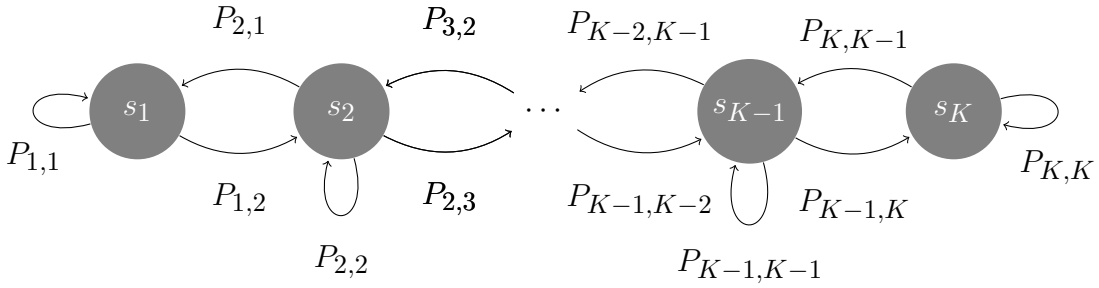


Figure A.1: State Transitions for FSMC Model of Channel SNR

it can be assumed that from frame to frame, the channel may only change to adjacent states. In this way, the transition probabilities are approximately given as in [19]:

$$P_{k,k+1} \approx \frac{N_\Lambda(\varphi_{k+1})T_f}{F_\Lambda(\varphi_{k+1}) - F_\Lambda(\varphi_k)} \quad k = 1, \ldots, K-1 \tag{A.18}$$

$$P_{k,k-1} \approx \frac{N_\Lambda(\varphi_k)T_f}{F_\Lambda(\varphi_{k+1}) - F_\Lambda(\varphi_k)} \quad k = 2, \ldots, K \tag{A.19}$$

$$P_{k,k} \approx 1 - P_{k,k-1} - P_{k,k+1} \quad k = 2, \ldots, K-1 \tag{A.20}$$

$$P_{1,1} \approx 1 - P_{1,2} \tag{A.21}$$

$$P_{K,K} \approx 1 - P_{K,K-1} \tag{A.22}$$

where $P_{m,n}$ denotes the probability of entering state n in the next frame, while the system is currently in state m, $N_\Lambda(\varphi)$ is the LCR of the SNR process in crossings per second at level φ and $F_\Lambda(\varphi)$ is the CDF of γ evaluated at φ. Figure A.1 shows the one dimensional state transition diagram describing the FSMC.

Appendix B
Optimization Techniques

Optimization theory plays an important role in a wide range of problem solving applications. The main purpose of this broad class of theories can be summarized follows.

> *To maximize or minimize an objective function subject to a number of constraints (linear or non-linear).*

With such a broad use, it is clear why optimization methods have found way into a broad range of fields ranging from economics, to control system theory.

Optimization algorithms can be characterized into two subfields of problem solving methods. Those classified as static or dynamic in nature.

B.1 Static Programming

Methods of solving static programming problems are resolved from the fundamental problem to solve. Problems solved using these methods focus on optimizing the overall objective function. Problems solved using static methods may contain linear or non-linear constraints. Solving methods depend on the type of constraints.

B.2 Linear Programming

The general minimization problem for linear systems takes the structures described by

$$\begin{aligned} \min_{\mathbf{x}} \quad & \mathbf{c}^T\mathbf{x} \\ \text{s.t.} \quad & \mathbf{A}\mathbf{x} \leq \mathbf{b} \text{ and } \mathbf{x} \geq 0 \end{aligned} \tag{B.1}$$

where $\mathbf{c}$ denotes the costs (or rewards) to maximize for each value of $\mathbf{x}$ and $\mathbf{A}$, $\mathbf{b}$ are the constraints of the linear system. There are numerous methods for solving these type of problems known as Linear Programming (LP) methods. Most famous is the Simplex Method.

Methods of solving LP problems normally require introduction of non-negative slack variables to convert the inequality to an equality. For example given a system with a single constraint $ax_1 \leq b, x \geq 0$, this constraint can be represented as

$$ax_1 + x_{s1} = b, \quad x_1, x_{s1} \geq 0 \tag{B.2}$$

where x_{s1} is denoted as a slack variable. Because linear programming algorithms solve constraint problems in this way, its easy to represent introduce a series of equality constraints into a system and therefore an additional constraint set can exist resulting in an extended linear problem given by

$$\begin{aligned} \min_{\mathbf{x}} \quad & \mathbf{c}^T\mathbf{x} \\ \text{s.t.} \quad & \mathbf{A}\mathbf{x} \leq \mathbf{b}, \mathbf{A}'\mathbf{x} = \mathbf{b}' \text{ and } \mathbf{x} \geq 0 \end{aligned} \tag{B.3}$$

B.3 Non-Linear Programming

Non-linear programming is a more general formulation of an optimization problems given as

$$\begin{aligned} \min_{\mathbf{x}} \quad & f(\mathbf{x}) \\ \text{s.t.} \quad & \mathbf{A}\mathbf{x} \leq \mathbf{b},\ \mathbf{A}_{eq}\mathbf{x}_{eq} = \mathbf{b}_{eq},\ \mathbf{x} \geq 0,\ \mathbf{c}(\mathbf{x}) \leq 0 \end{aligned} \tag{B.4}$$

where $\mathbf{A}$ and $\mathbf{A}_{eq}$ are matrices, $\mathbf{b}$ and $\mathbf{b}_{eq}$ are vectors, $\mathbf{c}(\mathbf{x})$ is a vector of non-linear functions evaluated at $\mathbf{x}$ and $f(\mathbf{x})$ is a scalar non-linear function evaluated at $\mathbf{x}$. There are numerous methods used to solve non-linear programming problems. Such methods include linearizing the problem, linear approximations, Lagrange multipliers, amongst others. Commercial software packages such as MATLAB [87], also incorporate a

combination of these methods to solve problems of the form given in (B.4).

B.4 Quadratic Programming

Quadratic programming (QP) is a special structure of NLP program with linear constraints and the cost function is in a quadratic form

$$\begin{aligned} \min_{\mathbf{x}} \quad & \frac{1}{2}\mathbf{x}^T\mathbf{Q}\mathbf{x} \\ \text{s.t.} \quad & \mathbf{A}\mathbf{x} \leq \mathbf{b},\ \mathbf{A}_{eq}\mathbf{x}_{eq} = \mathbf{b}_{eq},\ \mathbf{x} \geq 0 \end{aligned} \tag{B.5}$$

QP solvers are also built into packages such as MATLAB. It has been shown [88] that QP problems can be solved in polynomial time using the ellipsoid method if the matrix $\mathbf{Q}$ is positive definite.

B.5 Binary Programming

A binary programming (BP) problem is a linear programming problem where the variables can take on values of $\{0, 1\}$. The optimization problem is given in the general form used in this thesis is

$$\begin{aligned} \min_{\mathbf{x}} \quad & \mathbf{c}^T\mathbf{x} \\ \text{s.t.} \quad & \mathbf{A}\mathbf{x} \leq \mathbf{1}_N,\ \mathbf{A}_{eq}\mathbf{x} = \mathbf{1}_M, \quad x_j \in \{0, 1\}, \forall j \in \mathbf{x} \end{aligned} \tag{B.6}$$

where $\mathbf{c}$ is the real-valued vector containing the weighted power of choosing a given allocation, $\mathbf{x}$ is the vector of allocation selections, $\mathbf{A}_{eq}$ is a binary equality constraint matrix of M rows and $\mathbf{A}$ is a binary inequality constraint matrix of N rows. Well-known methods of solving BP problems include branch and bound and LP relaxation.

B.6 Dynamic Programming

Dynamic programming methods differ from traditional optimization methods as they focus on subsets of problems in an attempt to determine the optimal solution. More specifically, such problems are generally recursive in nature as solutions of individual subproblems are dependant.

B.7 Greedy Algorithm

A greedy algorithm is a dynamic method of solving optimization problems. The greedy algorithm is a well-known dynamic programming technique. In greedy based algorithms, optimization is performed on each subproblem in sequence in an attempt to solve the global optimization problem. By finding the locally optimal solutions, the greedy method may solve a large-scale, multi-stage optimization problem in low polynomial order.

There are no guarantees that greedy algorithms reach the globally optimal solution. However this trade-off is afforded in the use of a greedy algorithm as it provides speedy optimization to large-scale multi-step problems and may in some problems guarantee to provide a global optimization solution. One of the most famous greedy based algorithms offering the global optimal solution is Dijkstra algorithm for finding the shortest path [89]. In cases where the global optimal solution is not guaranteed, greedy algorithms may still be used as the tradeoff in computational complexity can outweigh the gap between the optimal and sub-optimal solution performance.

B.8 Constrained Markov Decision Process

MDP-based optimization provides a framework for solving optimization problems to maximize the total expected reward, given a reward is earned for each possible decision made during a given state and is a well-known [90] dynamic programming framework. MDP is described by a state space $\mathcal{S}$, an action space $\mathcal{A}(s)$, a transition probability matrix $\mathbf{P}(a, s)$ and a reward or cost $\mathcal{R}(a, s)$.

Constrained MDP (or C-MDP) theory is an extension to traditional MDP-based optimization framework in that optimization problems can be formulated with a series of constraints and corresponding state rewards. Like MDP, state transitions with C-MDP require the Markov Property.

The optimal solution to an MDP problem is policy μ^{opt} defining the actions taken in each system state such that the average expected reward is maximized. In C-MDP, the optimal policy is described as a marginal action density function $\mu^{opt}(a|s)$ denoting the probability of choosing action a given that the system is currently in state s. This density function is related to the steady-state probabilities as

$$\mu^{opt}(a|s) = \frac{\pi(a,s)}{\pi(s)} = \frac{\pi(a,s)}{\sum_{a \in \mathcal{A}(s)} \pi(a,s)} \tag{B.7}$$

where $\pi(a,s)$ is the steady-state probability of the system occupying state s and taking action a. Solving an C-MDP problem requires deriving the steady-state probability set $\pi(a,s), \forall a \in \mathcal{A}(s), \forall s \in \mathcal{S}$. In general, methods of solving these probabilities vary based on the problem construction. In [51], these are solved using general LP methods. In this way, the constraints, rewards and system properties may be described through a series of linear equalities/inequalities and through an objective function. More specifically recalling the general form of an LP problem

$$f(\mathbf{x}) = \arg\max_{\mathbf{x}} \mathbf{c}^T\mathbf{x}, \quad \mathbf{A}\mathbf{x} \leq \mathbf{b}, \mathbf{A}_{eq}\mathbf{x} = \mathbf{b}_{eq} \text{ and } \mathbf{x} \geq 0 \tag{B.8}$$

where $\mathbf{x}$ contains the steady-state probabilities $\pi(a,s)$ and $\mathbf{A}, \mathbf{A}_{eq}, \mathbf{b}, \mathbf{b}_{eq}$ contains the linear system constraints. The vector $\mathbf{c}$ contains the rewards associated with each action/state pair. Due to the structure of the objective function in the LP problem shown in B.8, $\mathbf{c}$ should contain negative valued rewards. In addition to problem specific constraints, the following constraints are inherent and must be considered when forming the optimization problem for C-MDP:

- Unity Property (the complete probability space):

$$\sum_{a \in \mathcal{A}(s), s \in \mathcal{S}} \pi(a, s) = 1 \tag{B.9}$$

- State Transition Balance Property (probability of entering a state must equal the probability being in that state)

$$\sum_{s \in \mathcal{S}} \sum_{a \in \mathcal{A}(s)} \pi(a, s) p^a_{s,s'} = \sum_{a' \in \mathcal{A}(s')} \pi(a', s'), \forall s' \in \mathcal{S} \tag{B.10}$$

- Non-Negativity Property (probability cannot be less than 0):

$$\pi(a, s) \geq 0, \forall a \in \mathcal{A}(s), \forall s \in \mathcal{S} \tag{B.11}$$

where $p^a_{s,s'}$ is the transition probability from state s to state s' given action a.

Appendix C
Proof of PER Approximation in Chapter 3

Lemma 2. *For a packet error rate that satisfies the following*

$$PER = 1 - \prod_{j=1}^{M} (1 - BER_j(\gamma))^{\alpha_j L},$$

where α_j denotes the portion of bits subject to $BER_j(\gamma)$ such that $\sum_j \alpha_j = 1$ and L denotes the number of bits per packet. The resulting packet error rate is approximately

$$PER \approx 1 - \left(1 - \sum_{j=1}^{M} \alpha_j BER_j(\gamma)\right)^{L}$$

Proof. Let $f(\gamma)$ denote $1 - PER$ such that

$$f(\gamma) = \prod_{j=1}^{M} (1 - BER_j(\gamma))^{\alpha_j L}$$

Taking the log of both sides we have

$$\begin{aligned} \log[f(\gamma)] &= \log\left[\prod_{j=1}^{M} (1 - BER_j(\gamma))^{\alpha_j L}\right] \\ &= \sum_{j=1}^{M} \alpha_j L \log[1 - BER_j(\gamma)] \end{aligned}$$

Since $BER_j(\gamma) << |1|$ in any region of interest and $\log(1 + x) \approx x$ near 0, we have

$$\log[f(\gamma)] \approx \sum_{j=1}^{M} -\alpha_j L \cdot BER_j(\gamma)$$

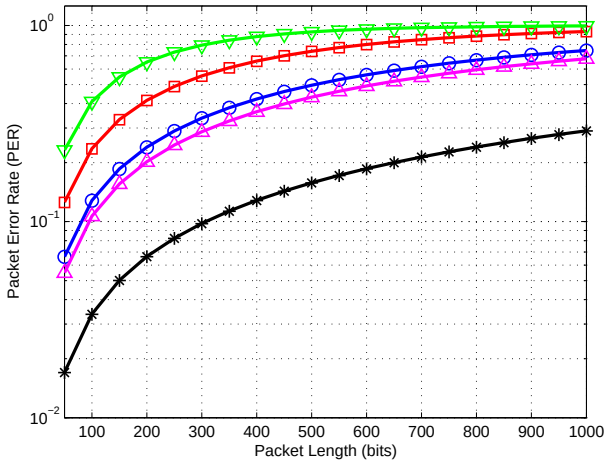


Figure C.1: PER approximation, 4 channels, BER vector $[10^{-5}\ 10^{-4}\ 10^{-3}\ 10^{-2}]$ and random bit assignments for α_j

By the same argument we note that $\alpha_j \leq 1$ by design and therefore $\sum_{j=1}^{M} -\alpha_j \cdot BER_j(\gamma) << |1|$ implying

$$\begin{aligned} \log[f(\gamma)] &\approx L \log\left[1 - \sum_{j=1}^{M} \alpha_j BER_j(\gamma)\right] \\ &\approx \log\left[\left(1 - \sum_{j=1}^{M} \alpha_j BER_j(\gamma)\right)^{L}\right] \end{aligned}$$

and by taking the exponential of both sides we have

$$\exp(\log[f(\gamma)]) = f(\gamma) \approx \left(1 - \sum_{j=1}^{M} \alpha_j BER_j(\gamma)\right)^{L}$$

thereby implying

$$PER \approx 1 - \left(1 - \sum_{j=1}^{M} \alpha_j BER_j(\gamma)\right)^{L}$$

In Figure C.1 one can observe the close approximation of the approximate PER to the actual PER (solid lines) for randomly generated sets of α_j for 4 channels. □

Appendix D
SC-FDMA Resource Allocation Optimal Gap

In order to address the performance of the iterative allocation algorithm described in Algorithm 2 used in Chapter 7, we compare its performance with the optimal allocation. The optimal optimizations formulation is as follows.

The formulation can be done using a similar approach as that used in [83] and in Chapter 6, however modified to operate for the MA resource allocation problem. In this fashion, the contiguity constraints are exploited in a manner that reduces the binary search space.

The optimization problem is solved at each subframe m. For brevity the index m is dropped however all quantities are assumed to be specific to subframe m. The problem can be expressed as a general set-packing problem and formulated using binary programming as

$$\begin{aligned} \min_{\mathbf{x}} \quad & \mathbf{c}^T\mathbf{x} \\ \text{s.t.} \quad & \mathbf{A}\mathbf{x} \leq \mathbf{1}_M,\ \mathbf{A}_{eq}\mathbf{x} = \mathbf{1}_K, \quad x_j \in \{0,1\}, \forall j \in \mathbf{x} \end{aligned} \tag{D.1}$$

where $\mathbf{c}$ is the real-valued vector containing the weighted power of choosing a given allocation, $\mathbf{x}$ is the vector of allocation selections, $\mathbf{A}_{eq}$ is a binary equality constraint matrix of K rows and $\mathbf{A}$ is a binary inequality constraint matrix of M rows. Each non-zero entry of the solution vector $\mathbf{x}$ corresponds to selecting the corresponding column allocation in $\mathbf{A}$.

The matrix $\mathbf{A}$ describes the set of potential RB allocations for all users. It is comprised of individual allocations given as

$$\mathbf{A} = [\mathbf{A}_1, \ \cdots, \ \mathbf{A}_K] \tag{D.2}$$

where $\mathbf{A}_i$ is a matrix containing the set of feasible allocations for UE i. Each column of $\mathbf{A}_i$ corresponds to a feasible allocation while each row corresponds to a specific resource. Each entry in $\mathbf{A}_i$ can take a value of $\{0, 1\}$. An entry of 1 if the particular resource is required by a UE for that allocation and 0 otherwise.

The set of possible allocation is determined as follows for each UE. During any subframe, a UE with data for transmission will utilize between 1 and M RBs in frequency. Any unique possible allocation is given as a column entry in $\mathbf{A}_i$. For example in the case where $M = 4$:

$$\mathbf{A}_i = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix} \tag{D.3}$$

The effective MCS scheme is a function of the number of RBs and T_i. For each possible contiguous block of RBs of size, the power level needed to maintain $BLER_{tgt}$ for all possible allocations of contiguous resource blocks is found using (7.5). For each such possible contiguous blocks above given by columns in $\mathbf{A}_i$, the corresponding transmission power is given in the corresponding entry of $\mathbf{c}$.

The equality matrix $\mathbf{A}_{eq}$ is simply a matrix of K rows constraining the number of selected allocations such that each UE is only allotted one allocation selection from their matrix $\mathbf{A}_i$. This is given as

$$\mathbf{A}_{eq} = \begin{bmatrix} \mathbf{1}_{C_1}^T & \cdots & \mathbf{0}_{C_K}^T \\ \vdots & \ddots & \vdots \\ \mathbf{0}_{C_1}^T & \cdots & \mathbf{1}_{C_K}^T \end{bmatrix} \tag{D.4}$$

where C_i is the number of columns in $\mathbf{A}_i$ and $\mathbf{1}_x$ and $\mathbf{0}_x$ are column vectors of length x.

The objective function vector $\mathbf{c}^T = [\mathbf{c}_1^T, \ldots, \mathbf{c}_K^T]$ is simply the cost of choosing the corresponding allocation for each $\mathbf{c}_i$. By the design of the problem, one can see the cost is simply the weighted power of choosing an allocation. Individual entries of $\mathbf{c}_i$ can be then be given as

$$c_{i:j_i} = \alpha_i P(\mathcal{N}_i(j_i), T_i), \quad j_i = 1, 2, \ldots, C_i \tag{D.5}$$

where the function $P_i(\cdot)$ is given in (7.5), α_i is the priority weight of UE i and where $\mathcal{N}_i(j_i) = \{x | a_{i:x,j_i} = 1, x = 1, 2, \ldots, M\}$. The quantity $a_{i:x,j_i}$ denotes the $\{x, j_i\}$ entry in $\mathbf{A}_i$ and j_i is the j_i^{th} column of $\mathbf{A}_i$.

In Figure D.1 and D.2 we show the result of the power allocation gap as a function of number users and resource blocks under the channel assumptions discussed in Chapter 7. Here we for example, all users to have a required data rate of 400 bits per frame with the same relative priority ($\alpha_i = 1, \forall i$). Figure D.2 is zoomed into a region of interest of Figure D.1. What we observe is under these conditions, that as long as the number of resource blocks is at least twice the number of users, the gap between the optimal and suboptimal allocation schemes is less than 15%. While this results in a relatively small increase in power allocated compared to the optimal allocation, there is relatively large reduction in computational complexity, and the latter method can be easily implemented in real-time. The holds true for TB sizes of interest presented in Chapter 7. For large TB sizes, the algorithm in Chapter 7 deviates from the optimal allocation. To improve its performance, this algorithm can be modified similar to that of Chapter 6 where the initial RB selection (lines 12–15 in Algorithm 2, Chapter 7) requires a minimum chunk allocation, (*i.e.,* $N_{min} > 1$ using notation developed in Chapter 6).

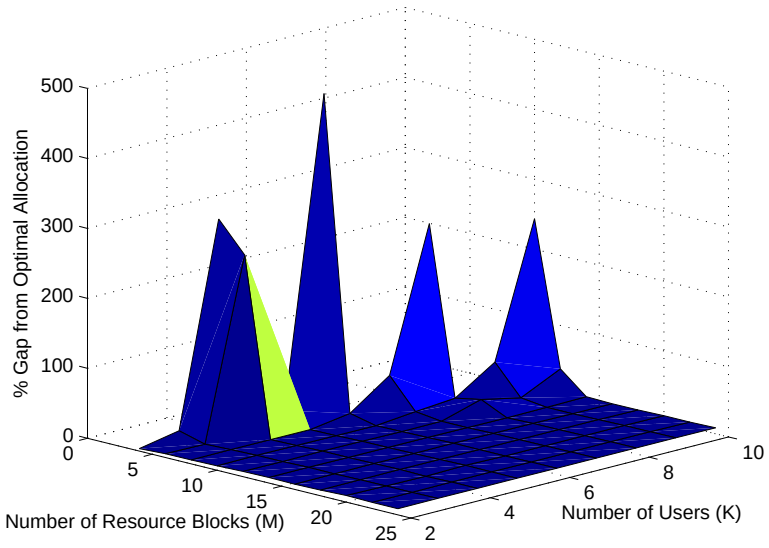


Figure D.1: Optimal/Suboptimal Allocation Gap

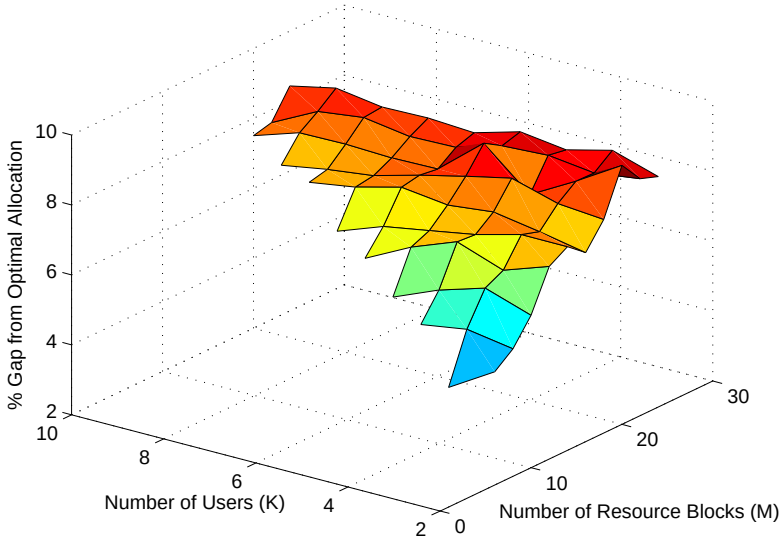


Figure D.2: Optimal/Suboptimal Allocation Gap - Zoom

Curriculum Vitae

Name: Dan J. Dechene

Post-secondary Education and Degrees:

2008-2011 Ph.D.
Communications Systems Engineering
The University of Western Ontario
London, Ontario, Canada

2006-2008 M.E.Sc.
Electrical and Computer Engineering
The University of Western Ontario
London, Ontario, Canada

2004-2006 B.Eng.
Electrical Engineering
Lakehead University
Thunder Bay, Ontario, Canada

2002-2004 Technologist
Electronics Engineering Technology
RCC Institute of Technology
Concord, Ontario, Canada

Refereed Journal Publications:

[J1] C. E. Kennedy, **D. J. Dechene**, and A. Shami, *Design, Implementation, and Evaluation of an FPGA-based WLAN Synchronizer.* In Press, Wiley Wireless Communications and Mobile Computing, 2011.

[J2] **D. J. Dechene** and A. Shami, *QoS, Channel and Energy-Aware Packet Scheduling over Multiple Channels.* IEEE Transactions on Wireless Communications, vol.10, no.4, pp.1058-1062, 2011.

[J3] **D. J. Dechene** and A. Shami, *Energy Efficient Quality of Service Traffic Scheduler for MIMO Downlink SVD Channels.* IEEE Transactions on Wireless Communications, vol.9, no.12, pp.3750-3761, 2010.

[J4] **D. J. Dechene** and A. Shami, *Performance Evaluation of MIMO-Aware Media Access Control Protocol.* Physical Communication, 2009, vol. 2, no. 3, pp. 204-216, 2009.

Refereed Conference Publications:

[C1] **D. J. Dechene** and A. Shami, "Energy Efficient Resource Allocation in SC-FDMA Uplink with Synchronous HARQ Constraints," in Proc of IEEE ICC'11 Wireless Networking 2011.

[C2] **D. J. Dechene**, S. Primak, and A. Shami, "On the First and Second Order Statistics of Sparse MIMO Channels," in Proc of 25th Biennial Symposium on Communications (QBSC), 2010, pages 391-394.

[C3] S. Haghighi, **D. J. Dechene**, A. Shami, S. Primak, and X. Wang, "On Energy Efficiency of Pilot Assisted Modulation Schemes," in Proc of IEEE Sarnoff Symposium, 2010.

[C4] **D. J. Dechene** and A. Shami, "Experimental Triple-Play Service Delivery Using Commodity Wireless LAN Hardware," In Proc. of IEEE ICC'09 Wireless Networking 2009.

[C5] P. Kumarawadu, **D. J. Dechene**, M. Luccini and A. Sauer, "Algorithms for Node Clustering in Wireless Sensor Networks: A Survey," In Proc. of IEEE ICIAFS 2008, pages 295-300.

[C6] **D. J. Dechene**, K. A. Meerja, A. Shami, and S. Primak, "A Novel MIMO-Aware Distributed Media Access Control Scheme for IEEE 802.11 Wireless Local Area Networks," In Proc. of IEEE LCN 2007, pages 125-132.

Teaching Assistantships:

[T1] ECE370/3370 - Communication Electronics
[T2] ECE436/4436 - Networking
[T3] ECE451 - Advanced Topics in Wireless Communications
[T4] SE314 - Wireless Communications and Networking
[T5] SE410/4410 - Wireless LANs and MANs

Awards:

[A1]	Ontario Graduate Scholarship (OGS)	**2011**
[A2]	IEEE Communications Society Travel Award	**2009**
[A3]	NSERC Post Graduate Scholarship D	**2008–2011**
[A4]	Best Presentation Award ECE Symposium	**2008 & 2010**
[A5]	Faculty of Engineering (ECE) Travel Award	**2007, 2009–2011**