

# **Towards Efficient and Fair Radio Resource Allocation Schemes for Interference-Limited Cellular Networks**

by

**Akram Bin Sediq**

A thesis submitted to the Faculty of Graduate and Postdoctoral  
Affairs in partial fulfillment of the requirements  
for the degree of

**Doctor of Philosophy**

in

Electrical and Computer Engineering

Carleton University  
Ottawa, Ontario, Canada  
September 2013

© 2013 - Akram Bin Sediq

The undersigned recommend to  
the Faculty of Graduate and Postdoctoral Affairs  
acceptance of the thesis

**Towards Efficient and Fair Radio Resource Allocation  
Schemes for Interference-Limited Cellular Networks**

Submitted by **Akram Bin Sediq**  
in partial fulfilment of the requirements for the degree of  
**Doctor of Philosophy**

---

Professor Halim Yanikömeroğlu, Thesis Supervisor

---

Professor Alagan Anpalagan, External Examiner  
Dept. of Electrical and Computer Engineering,  
Ryerson University, Toronto, Canada

---

Professor Roshdy Hafez, Department Chair

Carleton University

2013

# Abstract

The focus of this thesis is on studying the tradeoff between efficiency and fairness in interference-limited cellular networks. We start by characterizing the optimal tradeoff between efficiency and fairness in general resource allocation problems, including those encountered in cellular networks, where efficiency is measured by the sum-rate and fairness is measured by the Jain's fairness index. Among the commonly-used methods to approach these problems is the one based on the  $\alpha$ -fair policy. Analyzing this policy, we show that it does not necessarily achieve the optimal Efficiency-Jain Tradeoff (EJT) except for the case of two users. When the number of users is greater than two, we prove that the gap between the efficiency achieved by the  $\alpha$ -fair policy and that achieved by the optimal EJT policy for the same Jain's index can be unbounded. Finding the optimal EJT corresponds to solving potentially difficult non-convex optimization problems. To alleviate this difficulty, we derive sufficient conditions, which are shown to be sharp and naturally satisfied in various radio resource allocation problems. These conditions provide us with a means for identifying cases in which finding the optimal EJT can be reformulated as convex optimization problems. The new formulations are used to devise computationally-efficient resource schedulers that achieve the optimal EJT and surpass the baseline schedulers in terms of sum-rate efficiency, Jain's fairness index, median rate, and user satisfaction, without incurring additional complexity.

Applying the proposed optimal EJT schedulers in interference-limited cellular

networks requires the development of efficient interference management techniques. Inter-cell interference coordination (ICIC) is one of the most prominent techniques to manage interference by coordinating the allocation of radio resources across multiple cells. The problem of achieving various efficiency-fairness tradeoffs, including optimal EJT, in the long-term average rates can be cast as a multi-cell weighted sum-rate maximization optimization problem. By identifying a separable structure and a network-flow structure, we show that such optimization problem is amenable to powerful optimization methods, including the primal-decomposition method, the projected-subgradient method, and the network-flow optimization methods. Using these optimization methods, we propose a polynomial-time distributed ICIC scheme that finds a near-optimum multi-cell resource allocations. In comparison with the baseline static and dynamic ICIC schemes, the proposed scheme is shown to achieve higher gains in efficiency, Jain's fairness index, cell-edge rate, and outage probability.

Dedicated to Safia Zainal and Ahmed Zainal

# Acknowledgments

First and foremost, I am grateful to Allah (God), the Gracious and the Merciful, for all that I am and all that I have. To my parents, Salem and Ferdous, I cannot thank you enough for the unconditional love and support since my first breath in this world. To my lovely wife, Areej, I must express my sincere appreciation for providing the best environment a husband can hope for to do this research, despite being an undergraduate student yourself with a very busy schedule. To my sisters and brothers, thanks for all your support during this research. To Dr. Mohammed Ali Al-Ghalayini and Mr. Mohammed Gayed, I would like to express my sincere gratitude for mentoring and guiding me throughout my undergraduate and graduate studies.

I am very thankful for the encouragement, patient guidance, and the endless support I received from my supervisor Dr. Halim Yanıkömeroğlu during my Master's and PhD studies. Most notably, I would like to thank him for his tremendous efforts to build and continuously expand such a dynamic and large research group with members from around the globe. Needless to say, the interaction with the members of the research group has significantly contributed to my knowledge and skills.

I am also thankful to the defence committee members, Prof. Alagan Anpalagan, Prof. Minyi Huang, Prof. Abbas Yongacoglu, and Prof. Roshdy Hafez, for their valuable comments and discussions which improved the content and the presentation of this thesis.

Many thanks go to all the members of our wonderful research group for their encouragement during the course of this research. Most notably, I would like to thank my co-author and friend Dr. Ramy H. Gohary for patiently teaching me a great deal of optimization, technical writing, history, and politics. My thanks as well extend to my project managers, co-authors, and friends, Dr. Rainer Schoenen and Dr. Petar Djukic for their helpful comments and suggestions on my research.

To Muhammad Al-Juaid and Ziad El-Khatib, words can never thank you enough for the continuous support and assistance you provided me since my first day in Ottawa.

I would like also to thank the industry collaborator, Huawei Technologies, for funding in part the research in this thesis and filing three patent applications from my research. In addition, I would like to thank Dr. Gamini Senarath and Dr. Ho Ting Cheng from Huawei Technologies Canada for their contribution in putting our work into industrial perspective and for being instrumental in the patent filing process.

This work is also supported in part by the Ontario Graduate Scholarships (OGS) and in part by the Queen Elizabeth II Graduate Scholarships in Science and Technology (QEII-GSST); I thankfully acknowledge this support.

# Table of Contents

<b>Abstract</b>	<b>iii</b>
<b>Acknowledgments</b>	<b>vi</b>
<b>Table of Contents</b>	<b>viii</b>
<b>List of Tables</b>	<b>xiii</b>
<b>List of Figures</b>	<b>xiv</b>
<b>List of Acronyms</b>	<b>xviii</b>
<b>List of Mathematical Operators and Symbols</b>	<b>xx</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Motivation . . . . .	1
1.2 Overview . . . . .	3
1.2.1 Optimal Tradeoff between Sum-Rate Efficiency and Jain’s Fairness Index in Resource Allocation . . . . .	3
1.2.2 Inter-Cell Interference Coordination . . . . .	4
1.3 Contributions . . . . .	5
1.4 Organization . . . . .	6

## 2 Optimal Tradeoff between Sum-Rate Efficiency and Jain's Fairness

<b>Index in Resource Allocation: Theoretical Framework</b>	<b>7</b>
2.1 Introduction . . . . .	7
2.2 Preliminaries . . . . .	11
2.3 Does $\alpha$ -fair Policy Achieve the Optimal Efficiency-Jain Tradeoff? . . .	13
2.4 The Optimal Efficiency-Jain Tradeoff Policy . . . . .	18
2.4.1 A Technique for Obtaining the Optimal EJT for an Arbitrary $\mathcal{C}$	18
2.4.2 A Property for Ensuring Tractability . . . . .	23
2.4.3 Sufficient Conditions for Satisfying the Monotonic Tradeoff Property . . . . .	25
2.4.4 Geometric Interpretation of the Optimal EJT . . . . .	31
2.4.5 An Alternate Formulation . . . . .	33
2.5 Conclusions . . . . .	35

## 3 Optimal Tradeoff between Sum-Rate Efficiency and Jain's Fairness

<b>Index in Resource Allocation: Applications</b>	<b>37</b>
3.1 Introduction . . . . .	37
3.2 Congestion Control . . . . .	39
3.2.1 Special Case: Linear Networks . . . . .	39
3.3 Radio Resource Allocations in Quasi-Static Channels: Analysis . . . .	44
3.3.1 Special Case: Scheduling in Frequency-Flat Fading with Time- Sharing . . . . .	46
3.4 Radio Resource Allocations in Quasi-Static Channels: Numerical Results	51
3.4.1 Frequency-Flat Fading . . . . .	52
3.4.2 Frequency-Selective Fading . . . . .	54
3.5 Radio Resource Allocations in Ergodic Time-Varying Channels: Analysis	57
3.5.1 The Gradient Scheduling Algorithm . . . . .	58

3.5.2	Application of Gradient Scheduling to Efficiency-Fairness	
	Utilities . . . . .	60
3.6	Radio Resource Allocations in Ergodic Time-Varying Channels: Simulation Results . . . . .	64
3.6.1	Simulation Setup and Parameters . . . . .	64
3.6.2	Validation of the Simulator . . . . .	66
3.6.3	Comparison Results . . . . .	76
3.7	Conclusions . . . . .	82
<b>4</b>	<b>Optimized Distributed Inter-cell Interference Coordination Scheme using Projected Subgradient and Network Flow Optimization</b>	<b>83</b>
4.1	Introduction . . . . .	83
4.2	System Model . . . . .	88
4.3	Problem Statement . . . . .	90
	4.3.1 Power Control . . . . .	90
	4.3.2 Problem Formulation . . . . .	93
4.4	Proposed Algorithm . . . . .	95
	4.4.1 Bound Optimization . . . . .	96
	4.4.2 Transforming (4.13) Into an Equivalent Binary Linear Optimization Problem . . . . .	100
	4.4.3 Linear Optimization Relaxation . . . . .	103
	4.4.4 Primal Decomposition . . . . .	106
	4.4.5 Transforming (4.24) into an Equivalent Minimum Cost Network Flow Problem . . . . .	109
	4.4.6 Pseudocode . . . . .	114
	4.4.7 Intuitive Interpretation . . . . .	114
	4.4.8 Complexity Analysis of the Proposed Algorithm . . . . .	115

4.5	Baseline Schemes . . . . .	116
4.6	Simulation Results . . . . .	117
4.6.1	Optimality Gaps . . . . .	118
4.6.2	Comparing the Proposed Scheme with the Baseline Schemes . . . . .	120
4.6.3	Statistics of the Average Number of Restricted RBs . . . . .	129
4.7	Conclusions . . . . .	129
<b>5</b>	<b>Inter-cell Interference Coordination in Heterogeneous Networks</b>	<b>133</b>
5.1	Introduction . . . . .	133
5.2	ICIC schemes in Heterogeneous Networks . . . . .	136
5.2.1	Baseline schemes . . . . .	136
5.2.2	Proposed schemes . . . . .	137
5.2.3	Proposed Dynamic ABS Scheme . . . . .	139
5.3	Simulation Setup . . . . .	140
5.3.1	Wideband SINR CDF . . . . .	141
5.4	Simulation Results . . . . .	143
5.4.1	Fixed Number of Pico-BSs and Fixed Range Extension Margin . . . . .	143
5.4.2	Variable Number of Pico-BSs and Fixed Range Extension Margin . . . . .	147
5.5	Conclusions . . . . .	151
<b>6</b>	<b>Conclusions and Future Work</b>	<b>153</b>
6.1	Summary and Contributions . . . . .	153
6.2	Future Work . . . . .	156
<b>A</b>	<b>Published Papers, Submitted Manuscripts, and Patent Applications</b>	<b>159</b>
A.1	Patents . . . . .	159

A.2 Journal Papers . . . . .	160
A.3 Conference Papers . . . . .	161
<b>List of References</b>	<b>162</b>

## List of Tables

3.1	Simulation parameters based on IMT-Advanced UMa scenario. . . . .	67
3.2	The AMC table . . . . .	69
4.1	Mean and standard deviation of the optimality gap (%) . . . . .	119
5.1	Simulation parameters based on 3GPP heterogeneous network deployment Scenario 6.2. . . . .	142
5.2	Simulation parameters based on 3GPP heterogeneous network deployment Scenario 6.2 (cont'd). . . . .	144

# List of Figures

2.1	A graphical illustration of the definition of the optimal EJT provided in Definition 2. . . . .	13
2.2	The monotonic tradeoff property: An illustrative example. . . . .	30
2.3	An example of a set that does not possess the monotonic tradeoff property. . . . .	32
2.4	The optimal EJT benefit vector, $\mathbf{x}_\sigma^*$ , is the unique projection of the fairest vector $\frac{\sigma}{M}\mathbf{1}_M$ onto the set $\{\mathbf{x} \eta(\mathbf{x}) = \sigma, \mathbf{x} \in \mathcal{C}\}$ . Projection of $\beta\mathbf{1}_M$ onto $\mathcal{C}$ is also shown. . . . .	33
3.1	A linear network that consists of $L$ links and $L + 1$ users. User $m$ , $1 \leq m \leq L$ , uses only the $m^{\text{th}}$ link, while user $L + 1$ uses all the $L$ links. . . . .	39
3.2	A comparison between the EJTs achieved by the optimal and $\alpha$ -fair policies in a linear network. . . . .	43
3.3	A comparison between the EJTs achieved by the optimal and $\alpha$ -fair schedulers in frequency-flat channels with time-sharing. . . . .	54
3.4	A comparison between the users' rates achieved by the optimal EJT and $\alpha$ -fair schedulers in frequency-flat channels with time-sharing at an efficiency of 1 Mb/s. . . . .	55
3.5	A comparison between the EJTs achieved by the optimal and $\alpha$ -fair schedulers with time-sharing. . . . .	56
3.6	A comparison between the EJTs achieved by the optimal and $\alpha$ -fair schedulers without time-sharing. . . . .	57

3.7	A network layout that consists of 57 hexagonal sectors (without wrap around).	68
3.8	CDF of the wideband SINR for the network layout shown in Fig. 3.7.	68
3.9	A wrap around network layout that consists of a cluster of 57 hexagonal sectors (the middle cluster) and 6 “virtual” clusters.	71
3.10	CDF of the wideband SINR for the network layout shown in Fig. 3.9.	72
3.11	CDF of the wideband SINR for different association strategies.	74
3.12	CDF of the pathgain.	74
3.13	Distribution of the normalized time-average user throughput obtained using our simulator.	75
3.14	CDF of the normalized time-average user throughput produced by seven WINNER+ partners.	75
3.15	A long-term average rates comparison between the EJT's achieved by the $\alpha$ -fair, the MTMR, and the proposed optimal EJT schedulers.	77
3.16	Percentage of satisfied users for different thresholds.	79
3.17	A long-term average rates comparison of the Efficiency-Median Trade-off achieved by the $\alpha$ -fair, the MTMR, and the proposed optimal EJT schedulers.	80
3.18	A long-term average rates comparison of the Efficiency-Cell-edge Tradeoff achieved by the $\alpha$ -fair, the MTMR, and the proposed optimal EJT schedulers.	81
4.1	Illustrations of reuse-1 and reuse-3.	84
4.2	Illustration of PFR.	85
4.3	The considered network layout which consists of 19 cell sites and 3 sectors per site.	88
4.4	A block diagram that shows the steps used to develop the proposed algorithm.	97

4.5	Equivalent graphical representation of the minimum cost network flow problem given by (4.31). Vertices are labeled according to the label of their corresponding constraints. All flows are restricted to be between 0 and 1. . . . .	113
4.6	CDF of the normalized user throughput of all UTs in the network for $\alpha = 2$ and $\bar{R}_{\min} = 0$ . . . . .	121
4.7	Normalized cell-edge throughput versus normalized efficiency, for $\alpha = 1$ and $\bar{R}_{\min} \in \{0, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07\}$ . . . . .	122
4.8	Normalized cell-edge throughput versus normalized efficiency for different schemes, for $\bar{R}_{\min} = 0$ and $\alpha \in \{0, 0.25, 0.50, 0.75, 1, 2, 3, 4, 5\}$ . . . . .	123
4.9	Gain (%) in normalized cell-edge throughput versus normalized efficiency, as compared to reuse-1. . . . .	124
4.10	Gain (%) in normalized efficiency versus normalized cell-edge throughput, as compared to reuse-1. . . . .	124
4.11	Network layout with nonuniform user distributions, where each gray-sector has 10 UTs and each white-sector has 2 UTs. . . . .	125
4.12	Jain's index versus normalized efficiency, for $\beta \in \{0.03, 0.04, 0.05, 0.075, 0.1, 0.15, 0.2, 0.3, 0.5, 0.8, 1, 2, 5\}$ (bit/sec/Hz) . . . . .	127
4.13	Gain (%) in normalized efficiency versus Jain's index, as compared to reuse-1. . . . .	128
4.14	Gain (%) in Jain's index versus normalized efficiency, as compared to reuse-1. . . . .	130
4.15	Outage probability versus $\bar{R}_{\min}$ , which is defined as the probability that the average UT throughput is less than $\bar{R}_{\min}$ . . . . .	131
4.16	Probability mass function of the average number of restricted RBs per sector (the total number of RBs is 50). . . . .	131
5.1	Illustration of a HetNet. . . . .	134

5.2	CDF of the wideband SINR of macro-UTs and pico-UTs. . . . .	143
5.3	Gain (%) in efficiency of macro-UTs and pico-UTs, as compared to reuse-1. . . . .	145
5.4	Gain (%) in cell-edge throughput for macro-UTs and pico-UTs, as compared to reuse-1. . . . .	146
5.5	Gain (%) in median throughput for macro-UTs and pico-UTs, as compared to reuse-1. . . . .	147
5.6	Gain (%) in Jain's index for macro-UTs and pico-UTs, as compared to reuse-1. . . . .	148
5.7	Gain (%) in efficiency of all UTs as compared to reuse-1 for different number of pico-BSs. . . . .	149
5.8	Gain (%) in cell-edge throughput for all UTs as compared to reuse-1 for different number of pico-BSs. . . . .	150
5.9	Gain (%) in median throughput for all UTs as compared to reuse-1 for different number of pico-BSs. . . . .	151
5.10	Gain (%) in Jain's fairness index for all UTs as compared to reuse-1 for different number of pico-BSs. . . . .	152

# List of Acronyms

---

<b>Acronym</b>	<b>Definition</b>
ABS	Almost-Blank-Subframe
AMC	Adaptive Modulation and Coding
BS	Base-Station
CDF	Cumulative Distribution Function
DSL	Digital Subscriber Line
EJT	Efficiency-Jain Tradeoff
FFR	Fractional Frequency Reuse
HetNet	Heterogeneous Network
ICIC	Inter-Cell Interference Coordination
IMT-Advanced	International Mobile Telecommunications-Advanced
KKT	Karush-Kuhn-Tucker
LOS	Line-Of-Sight
LTE	Long-Term Evolution

<b>Acronym</b>	<b>Definition</b>
MTMR	Maximum-Throughput with Minimum-Rate
NLOS	Non-Line-Of-Sight
OFDM	Orthogonal Frequency Division Multiplexing
OFDMA	Orthogonal Frequency Division Multiple Access
PFR	Partial Frequency Reuse
RB	Resource Block
SFR	Soft Frequency Reuse
SINR	Signal-to-Interference-plus-Noise Ratio
SON	Self-Organizing Networks
UMa	Urban Macro
UT	User Terminal
WiMax	Worldwide Interoperability for Microwave Access

# List of Mathematical Operators and Symbols

---

Symbol	Definition
$\mathbb{R}_+^M$	The set of length- $M$ vectors with non-negative real entries
$\mathbf{1}_M$	The length- $M$ all-one vector
$\mathbf{0}_M$	The length- $M$ all-zero vector
$\succeq, \preceq$	Element-wise inequalities
$(\cdot)^T$	The transpose operator
$\ \cdot\ $	The Euclidean norm
$\mathbf{x}$	A vector of non-negative real entries $\{x_m\}_{m=1}^M$ , where $x_m$ is the benefit (e.g., rate) received by user $m$
$\mathcal{C}$	The set of feasible benefit vectors
$\eta(\mathbf{x})$	The efficiency that corresponds to benefit vector $\mathbf{x}$ which is defined as $\eta(\mathbf{x}) = \sum_{m=1}^M x_m$
$J(\mathbf{x})$	The Jain's fairness index that corresponds to benefit vector $\mathbf{x}$ which is defined as $J(\mathbf{x}) = \left(\sum_{m=1}^M x_m\right)^2 / M \sum_{m=1}^M x_m^2$

---

Symbol	Definition
--------	------------

---

$U_\alpha(\cdot)$	$\alpha$ -fair utility
$U_\beta(\cdot)$	Optimal Efficiency-Jain Tradeoff (EJT) utility
$M$	Number of users per sector, assuming uniform user distribution
$M^{(k)}$	Number of users in sector $k$
$N$	Number of available resource blocks (RBs)
$K$	Number of sectors
$L$	Number of finite-capacity links in a network carrying elastic-traffic
$\mathcal{M}^{(k)}$	Set of indices of users in sector $k$ , i.e., $\mathcal{M}^{(k)} \triangleq \{1, \dots, M^{(k)}\}$
$\mathcal{N}$	Set of indices of available RBs, i.e., $\mathcal{N} \triangleq \{1, \dots, N\}$
$\mathcal{K}$	Set of indices of all sectors in the network, i.e., $\mathcal{K} \triangleq \{1, \dots, K\}$
$\mathcal{K}^{(k)}$	Set of indices of the 6 first-tier interfering sectors seen by sector $k$
$m$	UT index; $m \in \mathcal{M}^{(k)}$
$n$	RB index; $n \in \mathcal{N}$
$k$	Sector index; $k \in \mathcal{K}$

---

Symbol	Definition
$H_{m,n}^{(k,\tilde{k})}$	Channel gain from sector $\tilde{k}$ on RB $n$ to UT $m$ served by sector $k$
$R_{m,n}^{(k)}$	Achievable rate on RB $n$ for UT $m$ in sector $k$
$r_{m,n}^{(k)}$	Achievable rate on RB $n$ for UT $m$ in sector $k$ , if RB $n$ is used by all other interfering sectors
$\tilde{r}_{m,n}^{(k,\tilde{k})}$	Additional rate on RB $n$ for UT $m$ in sector $k$ , gained if RB $n$ is restricted in sector $\tilde{k}$
$\bar{R}_m^{(k)}$	Average rate for UT $m$ in sector $k$
$\Gamma_{m,n}^{(k)}$	SINR on RB $n$ of UT $m$ in sector $k$
$\gamma_{m,n}^{(k)}$	SINR on RB $n$ of UT $m$ in sector $k$ , if all sectors use RB $n$
$\tilde{\gamma}_{m,n}^{(k,\tilde{k})}$	SINR on RB $n$ of UT $m$ in sector $k$ , if sector $\tilde{k}$ is restricted from using RB $n$
$\rho_{m,n}^{(k)}$	Binary variable indicating whether RB $n$ is assigned to UT $m$ in sector $k$ or not
$I_n^{(k)}$	Binary variable indicating whether RB $n$ is restricted in sector $k$ or not

# Chapter 1

## Introduction

### 1.1 Motivation

The ever-increasing demand for high-rate mobile data access is expected to grow at an unprecedented pace. For instance, according to a recent mobile traffic forecast report by Cisco [1], global mobile data traffic is expected to increase 18-fold by 2016 as compared to 2011. Such a substantial increase in mobile data traffic presents a tremendous challenge to service providers due to mainly the scarcity of radio resources.

Allocating scarce radio resources to multiple users encounters conflicting goals between efficiency and fairness. For instance, favouring a certain class of users may increase the system efficiency, but would result in the dissatisfaction of other classes of users. In contrast, providing equal benefits to all users may result in higher fairness, but will potentially result in low efficiency. In the context of radio resource allocations considered in this thesis, efficiency is measured by the sum-rate delivered to the users of the network, while fairness is measured by the widely used Jain's fairness index [2]—a bounded continuous measure that conforms to standard fairness benchmarks. To control the emphasis placed on either fairness or efficiency aspects, the provider uses a tradeoff policy, which, unless properly chosen, results in wasteful allocations. Hence,

for efficient utilization of resources, it is crucial to design an optimal Efficiency-Jain Tradeoff (EJT) policy whereby the system provider allocates its resources in such a way that no other allocation provides a strictly higher sum-rate efficiency and at the same time be fairer to the users, as measured by Jain's index.

In order to meet the ever-increasing demand for high-rate mobile data access, we need not only to efficiently and fairly allocate the scarce resources, but also to aggressively reuse these resources across the network. However, uncontrolled aggressive reuse can potentially result in unacceptable interference levels, which can complicate the allocation of these resources in a fair and an efficient manner. For instance, reuse-1, in which all radio resources are used in every cell, is an example of an aggressive frequency reuse scheme. While reuse-1 can potentially achieve high sum-rate efficiency, it can result in unacceptable fairness levels as it jeopardizes the throughput experienced by users close to the cell-edge, due to the excessive interference experienced by these users. Conventionally, interference is mitigated by static resource partitioning and frequency/cell-planning, where cells within close proximity are assigned orthogonal resources. Static resource partitioning techniques suffer from two drawbacks. First of all, the sum-rate efficiency is significantly reduced since each cell is allocated only a fraction of the available resources. Secondly, frequency/cell-planning may not be possible in emerging cellular networks where new low power base-stations (BSs), such as femto-/pico-BSs, are expected to be deployed in an ad hoc manner, without prior planning. To overcome these drawbacks, inter-cell interference coordination (ICIC) is introduced as one of the most prominent techniques to manage interference by coordinating the allocation of radio resources across multiple cells. Such coordination can potentially improve the efficiency-fairness tradeoff in the entire network.

The main focus of this thesis is on investigating the optimal tradeoff between sum-rate efficiency and Jain's fairness index in resource allocation and devising distributed

ICIC schemes that can achieve such optimal tradeoff in interference-limited cellular networks.

## 1.2 Overview

### 1.2.1 Optimal Tradeoff between Sum-Rate Efficiency and Jain's Fairness Index in Resource Allocation

In the first part of the thesis, we study the tradeoff between the sum-rate efficiency and Jain's fairness index in general resource allocation problems. Such problems are frequently encountered in wireless communication systems with  $M$  users. Among the commonly-used methods to approach these problems is the one based on the  $\alpha$ -fair policy [3]. Analyzing this policy, we prove that it does not necessarily achieve the optimal EJT except for the case of  $M = 2$  users. When the number of users  $M > 2$ , it is shown that the gap between the efficiency achieved by the  $\alpha$ -fair policy and that achieved by the optimal EJT policy for the same Jain's index can be unbounded.

Many applications, including wireless communications ones, involve the tradeoff between the benefits (e.g., the rates) received by more than two users. Since in these cases, the  $\alpha$ -fair policy does not necessarily achieve the optimal EJT, we develop another technique for achieving this tradeoff.

Finding the optimal EJT corresponds to solving a family of potentially difficult non-convex optimization problems. To alleviate this difficulty, we derive sufficient conditions, which are shown to be sharp and naturally satisfied in various radio resource allocation problems. These conditions provide us with a means for identifying cases in which finding the optimal EJT can be reformulated as convex optimization problems. The new formulations are used to devise computationally-efficient resource

schedulers that enable the optimal EJT to be achieved for both quasi-static and ergodic time-varying communication scenarios.

Through extensive simulations, we demonstrate the superiority of the proposed optimal EJT scheduler as compared to the baseline schedulers in terms of sum-rate efficiency, Jain's fairness index, median rate, and user satisfaction.

In addition to the practical aspects of our contributions, we also cast new conceptual understandings of the optimal EJT. In particular, we provide intuitive geometric interpretations and physical meanings of optimal EJT. Moreover, we demonstrate that the tradeoff between efficiency and Jain's fairness index may not necessarily exist in some scenarios. In particular, we show that there exists scenarios where increasing efficiency leads to an increase in Jain's fairness index.

### **1.2.2 Inter-Cell Interference Coordination**

Most of the currently available ICIC schemes in the literature have one (or more) of the following drawbacks: requirement of a centralized controller, requirement of the rate adaptation function to be differentiable, or heuristic-based schemes with no comparison with the optimum scheme. The second part of the thesis is devoted to developing a novel ICIC scheme that overcomes these drawbacks and achieves the optimal EJT in interference-limited cellular networks.

The problem of achieving various efficiency-fairness tradeoff, including optimal EJT, in the long-term average rates can be cast as a weighted sum-rate maximization optimization problem. Unfortunately, this optimization problem can be shown to be strongly NP-hard optimization problem, i.e., it is not only hard to solve but also hard to approximate with guaranteed optimality gap. However, rather than pursuing a heuristic approach, we start by exploiting the characteristics of dominant interferers to carefully approximate the problem into a bound optimization problem. The resulted bound optimization problem belongs to the class of combinatorial optimization

problems which are not easy to solve in general; thus, we perform several mathematical transformations to the bound optimization problem in order to make it amenable to existing optimization methods. In particular, we realize that the bound optimization problem has a minimum cost network flow structure. This is an important insight to the ICIC optimization problem because by exploiting such structure, we can use network-based algorithms which have significantly reduced complexity as compared to the general-purpose convex or linear optimization algorithms [4, p. 402]. This is a key insight that allows us to devise a computationally-efficient polynomial-time distributed ICIC scheme. Although the scheme is developed solely based on mathematical derivations, it has simple and intuitive interpretation which would shed light on how distributed ICIC schemes can be designed. Simulation results of a small network show that the proposed scheme achieves a performance that is close to the optimal one achieved by exhaustive search, which validates our approximation and mathematical transformations. Furthermore, extensive simulation results show that the proposed scheme achieves significantly higher cell-edge throughput, higher aggregate throughput, and lower outage probability as compared to the baseline static and dynamic schemes.

Although the proposed scheme is mainly developed in the context of a homogeneous network, where the access network consists of macro-BSs, we demonstrate its effectiveness in the context of a heterogeneous network, where the access network consists of macro-BSs, pico-BSs, femto-BSs, and relays.

### 1.3 Contributions

The key high-level contributions of this thesis are the followings:

- Development of a general mathematical framework which enables us to identify practical scenarios for which the optimal EJT can be efficiently computed. The

developed framework is shown to be applicable to various resource allocation problems in communication networks. By applying this framework to the problem of scheduling radio resources in cellular networks, we devise novel resource schedulers that achieve the optimal EJT for both quasi-static and time-varying channels. Substantial gains are observed with respect to the baseline schedulers, without incurring additional computational complexity.

- Development of a novel ICIC scheme that runs in polynomial time and finds near-optimum dynamic multi-cell resource allocations that maximize a weighted sum-rate in the entire network in a distributed manner. Substantial gains are observed with respect to the baseline ICIC schemes in both homogeneous networks and heterogeneous networks.

## 1.4 Organization

The remainder of this thesis is organized as follows. In Chapter 2, we review the relevant literature in the area of efficiency-fairness tradeoff and present a theoretical framework to characterize the optimal EJT in general resource allocation scenarios. We then provide applications of this theoretical framework in Chapter 3, with special emphasis on radio resource allocation in the downlink of cellular networks. In Chapter 4, we review the literature relevant to ICIC and develop a novel distributed and dynamic ICIC scheme in a homogeneous network, where the access network consists solely of macro-BSs. In Chapter 5, we develop variants of the proposed ICIC scheme that will prove useful in the context of a heterogeneous network, where the access network consists of macro-BSs and pico-BSs. We conclude in Chapter 6 by summarizing the major contributions of this thesis and highlighting a number of research areas for future work. A list of the publications as well as the patent applications produced during the PhD program is provided in Appendix A.

## Chapter 2

# Optimal Tradeoff between Sum-Rate Efficiency and Jain's Fairness Index in Resource Allocation: Theoretical Framework

© 2013 IEEE. Parts of this chapter are reprinted, with permission, from:  
A. Bin Sediq, R. H. Gohary, R. Schoenen, and H. Yanıkömeroğlu, "Optimal tradeoff between sum-rate efficiency and Jain's fairness index in resource allocation," *IEEE Transactions on Wireless Communications*, vol. 12, no. 7, pp. 3496–3509, July 2013.

## 2.1 Introduction

The resources available for wireless communication systems are usually scarce and shared among multiple users. The way in which these resources are allocated determines the efficiency of the system and the benefits received by its users. Since the service provider is interested in maximizing the efficiency of the system and the users are interested in maximizing their own benefits, the allocation of resources is

typically encountered by conflicting goals. For example, providing equal benefits to all users results in higher fairness, but will potentially result in low efficiency. On the other hand, favouring a certain class of users may increase the system efficiency, but would result in the dissatisfaction of other classes of users. To control the emphasis placed on various goals, the provider uses a tradeoff policy, which, unless properly chosen, can result in wasteful allocation of resources. In particular, a suboptimal tradeoff policy can be less efficient and, at the same time, less fair to the users [3,5,6].

The benefits received by the users in the downlink of a wireless communication system can be measured by the rates at which data is delivered to these users. These rates are controlled by appropriate allocation of radio resources at the transmitter. For instance, the transmitter may allocate its resources in such a way that maximizes the sum of the rates delivered to the users. This allocation favours users that are geographically closer to the transmitter, but “starves” farther users. Although more efficient from the system perspective, such an allocation is unfair to the users at less advantageous locations [7,8]. A fairer allocation is one in which the minimum rate received by the users is maximized [3,6,9]. However, this allocation can result in unacceptable system efficiency, i.e., low sum-rate. Hence, it is desirable to find an optimal tradeoff policy whereby the system provider allocates its resources in such a way that no other allocation can provide a strictly higher efficiency and at the same time be fairer to the users. The focus of this chapter is two folds: 1) to develop a technique for obtaining an efficiency-fairness tradeoff that is optimal in a specific sense; and 2) to derive sufficient conditions, which, when satisfied by the set of feasible benefits, lead to efficiently computable optimal tradeoff and benefit vectors.

The applications that we will consider throughout the thesis are derived from practical radio resource scheduling problems that arise in wireless communication systems operating over quasi-static and ergodic time-varying channels. However, our analysis applies to a broader class of frameworks, including social and economics

ones [5, 10, 11].

To study the tradeoff between efficiency and fairness, we note that efficiency is usually defined depending on the particular resource allocation problem considered. For instance, in the case of wireless networks considered herein, efficiency is measured by the sum-rate delivered to the users of the network. In contrast, several definitions are used to quantify fairness. In [5], axioms that include continuity and homogeneity, and subsequent features, are provided to obtain a general class of plausible fairness measures. Among the members of that class are the entropy-based index [5] and Jain's fairness index [2]. In addition to the axioms and features provided in [5], we identify two more features that commend the use of Jain's index as a fairness measure.

- Conformity to standard fairness benchmarks: A fairness measure with this feature can be related to easily conceivable benchmarks. For instance, a Jain's index of  $p/100$  can be regarded as the fairness index of an equivalent resource allocation in which  $p\%$  of the users receive equal non-zero benefits and the remaining  $(100 - p)\%$  receive zero benefits [2]. Analogous relations between other metrics and standard benchmarks are not readily available.
- Accommodating more users: A good fairness measure enables more users with specific benefit requirements to be accommodated in the system. The superiority of Jain's index in that respect will be illustrated by numerical comparisons hereinafter.

A common approach to trading off efficiency with fairness in wireless networks is to allocate the resources in a way that maximizes the sum-rate efficiency while ensuring that the minimum rates achieved by the users exceed some prescribed bounds, e.g., [6, 11]. Varying these bounds over the set of feasible rates provides a means for controlling fairness [6]. A similar approach is to allocate the resources such that the

sum-rate efficiency is maximized while satisfying a predetermined set of proportional-rate constraints which can be varied to control fairness [12]. Another approach is to allocate the resources in a way that maximizes a parametric utility, whereby one or multiple parameters are used to control the emphasis on efficiency and fairness. A commonly used policy is the  $\alpha$ -fair one (also known as the  $\alpha$ -fair utility) [3], wherein various settings of a parameter  $\alpha$  yield allocations that achieve popular efficiency-fairness tradeoffs. For instance, setting  $\alpha = 0$  yields maximum efficiency, setting  $\alpha = 1$  yields proportionally fair allocations [13], and setting  $\alpha = \infty$  yields allocations that are fair in the max-min sense [3]. Motivations for using the  $\alpha$ -fair policy are provided in [5]. Generally speaking, increasing  $\alpha$  results in allocations that are fairer [5] in a sense that does not necessarily conform to Jain's index, as will be shown hereinafter. Other parametric utilities for trading off efficiency and fairness are considered in [14] and [15], and a comparison between multiple tradeoff criteria is provided in [16].

Compared with other measures, Jain's index provides a fairness criterion that takes into consideration all the users of the system, not only those users that are assigned minimal resources [2]. Maximizing Jain's index without wasting valuable resources requires optimal tradeoff between efficiency and this index. A question that arises is whether maximizing the well-studied  $\alpha$ -fair policy yields such an optimal tradeoff. To address this question, we show that  $\alpha$ -fair allocations are not guaranteed to achieve the optimal EJT except for the case of  $M = 2$  users. To overcome this drawback, we develop a generic technique for obtaining optimal EJT allocations. Unfortunately, this technique involves solving a family of potentially difficult non-convex optimization problems. To alleviate this difficulty, we derive sharp sufficient conditions which provide us with a means for identifying cases in which finding the optimal EJT can be reformulated as convex optimization problems. Interestingly, these sufficient conditions are satisfied by a wide range of resource allocation problems

in communications networks, as we will demonstrate in Chapter 3.

**Notation** Bold-face and regular-face fonts will be used to denote vectors and scalars, respectively. The set of length- $M$  vectors with non-negative real entries will be denoted by  $\mathbb{R}_+^M$  and the length- $M$  all-one and all-zero vectors will be denoted by  $\mathbf{1}_M$  and  $\mathbf{0}_M$ , respectively. The symbols  $\preceq$  and  $\succeq$  will be used to denote element-wise inequalities, and  $(\cdot)^T$  will be used to denote the transpose. The Euclidean norm will be denoted by  $\|\cdot\|$ .

## 2.2 Preliminaries

Let  $\mathbf{x} \in \mathcal{C} \subseteq \mathbb{R}_+^M$  be a vector of non-negative real entries  $\{x_m\}_{m=1}^M$ , where  $x_m$  is the benefit received by user  $m$  and  $\mathcal{C}$  is the set of feasible benefit vectors. Generally, the benefits  $\{x_m\}_{m=1}^M$  and the set  $\mathcal{C}$  depend on the application and the resources allocated to each user [2, Sec. 5]. For example, in the downlink of wireless communications,  $x_m$  can be the rate of user  $m$  resulting from a particular allocation of the radio resources, and  $\mathcal{C}$  can be the set of all achievable rates. In this thesis, the efficiency,  $\eta(\mathbf{x})$ , of a resource allocation is defined by the sum of benefits (i.e.,  $\eta(\mathbf{x}) = \sum_{m=1}^M x_m$ ), and its fairness is given by the Jain's index defined below.

**Definition 1** (Jain's Index [2]). *For  $\mathbf{x} \in \mathbb{R}_+^M$ , Jain's fairness index  $J : \mathbb{R}_+^M \rightarrow \mathbb{R}_+$  is given by*

$$J(\mathbf{x}) = \left( \sum_{m=1}^M x_m \right)^2 / M \sum_{m=1}^M x_m^2. \quad (2.1)$$

□

This definition shows that  $J(\mathbf{x})$  is continuous and lies in  $[\frac{1}{M}, 1]$ . In this interval,  $J = \frac{1}{M}$  corresponds to the least fair allocation in which only one user receives a non-zero benefit, and  $J = 1$  corresponds to the fairest allocation in which all users receive

the same benefit.

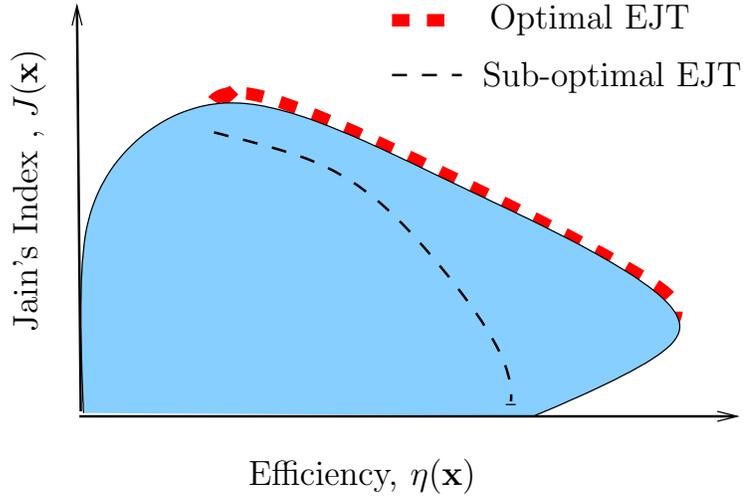
In many cases, depending on  $\mathcal{C}$ , there is an inherent tradeoff between  $\eta(\mathbf{x})$  and  $J(\mathbf{x})$ . Hence, to ensure efficient utilization of resources, we seek the optimal tradeoff, which is defined next.

**Definition 2** (Optimal EJT). *An optimal EJT is one that results in a benefit vector  $\mathbf{x}^*$  such that no  $\mathbf{x} \neq \mathbf{x}^*$ ,  $\mathbf{x} \in \mathcal{C}$  satisfies either: 1)  $\eta(\mathbf{x}) > \eta(\mathbf{x}^*)$ , and at the same time,  $J(\mathbf{x}) \geq J(\mathbf{x}^*)$ , or 2)  $\eta(\mathbf{x}) \geq \eta(\mathbf{x}^*)$ , and at the same time,  $J(\mathbf{x}) > J(\mathbf{x}^*)$ .  $\square$*

This definition is closely related to Pareto optimality defined for optimization problems with multiple objectives [17]. With efficiency and Jain's index as objectives, a Pareto optimal point is one at which efficiency cannot be increased without decreasing Jain's index and likewise, Jain's index cannot be increased without decreasing efficiency. As such, a point that is optimal from the EJT perspective, as per Definition 2, is equivalent to Pareto optimality in efficiency and Jain's index. However, a point that is Pareto optimal from an Efficiency-Jain perspective is not necessarily Pareto optimal if the multiple objectives are taken to be the users' benefits themselves, rather than the efficiency and Jain's index that these benefits achieve.

A graphical example that illustrates the definition of the optimal EJT is provided in Fig. 2.1. In this figure, the shaded area represents the feasible set of efficiencies and Jain's indexes for this example. The corresponding optimal EJT curve as well as a sub-optimal EJT curve are also shown in the figure. The sub-optimality of every point in the sub-optimal EJT curve can be readily verified by noting that there are feasible points to the north-east of this curve with simultaneously higher efficiency and higher Jain's index. On the contrary, there are no feasible points to the north-east of the optimal EJT curve which verifies its optimality from EJT perspective.

Definition 2 will be used in the next section to determine whether the  $\alpha$ -fair



**Figure 2.1:** A graphical illustration of the definition of the optimal EJT provided in Definition 2.

tradeoff policy achieves the optimal EJT.

## 2.3 Does $\alpha$ -fair Policy Achieve the Optimal Efficiency-Jain Tradeoff?

Given an  $\alpha \in [0, \infty)$ , the benefit vector  $\mathbf{x}_\alpha^*$  generated by the  $\alpha$ -fair tradeoff policy maximizes the  $\alpha$ -fair utility [3], i.e.,

$$\mathbf{x}_\alpha^* = \arg \max_{\mathbf{x} \in \mathcal{C}} U_\alpha(\mathbf{x}), \quad (2.2)$$

where

$$U_\alpha(\mathbf{x}) = \begin{cases} \sum_{m=1}^M \log x_m, & \alpha = 1, \\ \frac{1}{1-\alpha} \sum_{m=1}^M x_m^{1-\alpha}, & \alpha \geq 0, \alpha \neq 1. \end{cases} \quad (2.3)$$

The  $\alpha$ -fair policy thus described was considered in [5]. It was shown therein that,

for  $\alpha \neq 1$ ,  $\mathbf{x}_\alpha^*$  generated by (2.2) is the same as that generated by

$$\mathbf{x}_\alpha^* = \arg \max_{\mathbf{x} \in \mathcal{C}} \left( \left| \frac{\alpha}{1-\alpha} \right| \mathcal{L}(H_\alpha(\mathbf{x})) + \mathcal{L}(\eta(\mathbf{x})) \right), \quad (2.4)$$

where  $\mathcal{L}(\cdot) \triangleq \text{sgn}(\cdot) \log(|\cdot|)$ , and

$$H_\alpha(\mathbf{x}) = \text{sgn}(1-\alpha) \sqrt[\alpha]{\sum_{m=1}^M \left( \frac{x_m}{\eta(\mathbf{x})} \right)^{1-\alpha}}. \quad (2.5)$$

This equivalent formulation of the  $\alpha$ -fair policy provides insight into the role of  $\alpha$ . In particular, it can be seen that  $\mathcal{L}(\cdot)$  is monotonically increasing and that, for any  $\alpha \neq 1$ ,  $H_\alpha(\mathbf{x})$  provides a homogeneous fairness measure [5]. Hence, it can be seen that increasing  $\alpha$  places more emphasis on fairness at the expense of efficiency.

Using the above observations, it was argued in [5] that solving (2.4) yields a benefit vector that achieves the optimal tradeoff between  $H_\alpha(\mathbf{x})$  and  $\eta(\mathbf{x})$ . Although this explanation offers a better understanding, it presents the fairness component of the  $\alpha$ -fair policy as being parameterized by  $\alpha$ . Hence, according to this explanation, varying  $\alpha$  not only controls the emphasis placed on fairness, but also changes the fairness measure itself.

In [5], the  $\alpha$ -fair utility given in (2.3) is generalized by proposing a two-parameters utility given as  $U_{\alpha,\lambda} = H_\alpha(\mathbf{x})(\eta(\mathbf{x})^{\frac{1}{\lambda}})^{\frac{1}{\alpha}}$ ,<sup>1</sup> where  $\alpha$  is used to choose the fairness measure and  $\lambda \in \mathbb{R}$  is used to control the emphasis on efficiency. Using this utility, one can obtain various optimal efficiency-fairness tradeoff benefit vectors for a particular fairness measure by maximizing  $U_{\alpha,\lambda}$  for a fixed  $\alpha$  and different values of  $\lambda$  (cf. [5, 18] for guidelines on choosing  $\alpha$  and  $\lambda$ ). Unfortunately, while  $U_{\alpha,\lambda}$  is log-concave for  $\alpha \in (0, 1)$  and  $\lambda \in [0, \frac{\alpha}{1-\alpha}]$  [18], it is neither concave nor log-concave for  $\alpha > 1$ , which

---

<sup>1</sup>It is also shown in [5] that  $U_{\alpha,\lambda}$  generalizes various fairness measures, including Jain's fairness index, min-ratio, max-ratio, and entropy. Unfortunately, maximizing  $U_{\alpha,\lambda}$  can be potentially difficult due to the non-concavity of  $U_{\alpha,\lambda}$  for general  $\alpha$  and  $\lambda$ .

can potentially complicate the search for vectors that maximize  $U_{\alpha,\lambda}$  for  $\alpha > 1$ .

A question that arises is whether the  $\alpha$ -fair policy achieves the optimal efficiency-fairness tradeoff in practical resource allocation scenarios wherein the fairness measure does not depend on extrinsic parameters like  $\alpha$ . To address this question, in this section we investigate the relationship between the  $\alpha$ -fair policy and the optimal EJT. We begin by studying the case of  $M = 2$  users. The main result in this case is stated in the following proposition.

**Proposition 1.** *Let  $\mathcal{C}$  be an arbitrary set, possibly discrete, and let  $M = 2$ . For any  $\alpha \in (0, \infty)$ , the  $\alpha$ -fair benefit vector  $\mathbf{x}_\alpha^*$  generated by (2.2) achieves the optimal EJT.*

*Proof.* We will proceed by contradiction. Let  $\alpha \in (0, \infty)$  be given and suppose that  $\mathbf{x}_\alpha^*$  does not achieve the optimal EJT, that is, there exists a non  $\alpha$ -fair optimal vector  $\mathbf{x}$  such that either: 1)  $\eta(\mathbf{x}) > \eta(\mathbf{x}_\alpha^*)$  and  $J(\mathbf{x}) \geq J(\mathbf{x}_\alpha^*)$ ; or 2)  $\eta(\mathbf{x}) \geq \eta(\mathbf{x}_\alpha^*)$  and  $J(\mathbf{x}) > J(\mathbf{x}_\alpha^*)$ . We will show that such a vector  $\mathbf{x}$  results in  $U_\alpha(\mathbf{x}) > U_\alpha(\mathbf{x}_\alpha^*)$ , which contradicts the definition of  $\alpha$ -fair benefit vectors; cf. (2.2). We will focus on the first case. The proof for the second case follows similar lines and is omitted for brevity.

Since  $M = 2$ , we can define a parameter  $\omega = \frac{\max \mathbf{x}}{\min \mathbf{x}}$ . Using this  $\omega$ , we have  $J(\mathbf{x}) = \frac{(1+\omega)^2}{2(1+\omega^2)}$ . Now,  $\frac{dJ}{d\omega} = -\frac{\omega^2-1}{(\omega^2+1)^2}$ . Since, by definition,  $\omega \geq 1$ , it can be seen that  $J$  is monotonically decreasing in  $\omega$ . This with the fact that, in the considered case,  $J(\mathbf{x}) \geq J(\mathbf{x}_\alpha^*)$  implies that

$$\frac{\max \mathbf{x}}{\min \mathbf{x}} \leq \frac{\max \mathbf{x}_\alpha^*}{\min \mathbf{x}_\alpha^*}. \quad (2.6)$$

Since in this case we also have  $\eta(\mathbf{x}) > \eta(\mathbf{x}_\alpha^*)$ , it follows that  $\min \mathbf{x} + \max \mathbf{x} > \min \mathbf{x}_\alpha^* + \max \mathbf{x}_\alpha^*$ , which is equivalent to  $(1 + \frac{\max \mathbf{x}}{\min \mathbf{x}}) \min \mathbf{x} > (1 + \frac{\max \mathbf{x}_\alpha^*}{\min \mathbf{x}_\alpha^*}) \min \mathbf{x}_\alpha^*$ .

This inequality implies that

$$\min \mathbf{x} > \frac{\left(1 + \frac{\max \mathbf{x}_\alpha^*}{\min \mathbf{x}_\alpha^*}\right)}{\left(1 + \frac{\max \mathbf{x}}{\min \mathbf{x}}\right)} \min \mathbf{x}_\alpha^*. \quad (2.7)$$

Invoking (2.6) implies that the fraction on the right hand side is greater than 1, which further implies that we can write  $\min \mathbf{x} = \min \mathbf{x}_\alpha^* + \epsilon_1$ , with  $\epsilon_1 > 0$ . Since  $\mathbf{x}$  is not  $\alpha$ -fair, we must have

$$U_\alpha(\mathbf{x}) < U_\alpha(\mathbf{x}_\alpha^*). \quad (2.8)$$

We now observe that  $U_\alpha(\mathbf{x})$  is strictly increasing in each  $x_m$ ,  $m = 1, 2$ . This observation and (2.8) imply that  $\max \mathbf{x} = \max \mathbf{x}_\alpha^* - \epsilon_2$ , with  $\epsilon_2 > 0$ . Combining this with the fact that  $\min \mathbf{x} = \min \mathbf{x}_\alpha^* + \epsilon_1$  and the fact that in the current case  $\eta(\mathbf{x}) > \eta(\mathbf{x}_\alpha^*)$  yields  $\epsilon_1 > \epsilon_2$ . Using this notation, it can be readily verified that, because  $\epsilon_1 > \epsilon_2$ ,

$$\nabla U_\alpha(\mathbf{x})^T (\mathbf{x}_\alpha^* - \mathbf{x}) = -\epsilon_1 (\min \mathbf{x})^{-\alpha} \left( 1 - \frac{\epsilon_2}{\epsilon_1} \left( \frac{\max \mathbf{x}}{\min \mathbf{x}} \right)^{-\alpha} \right) < 0.$$

Now, direct computation of the Hessian of  $U_\alpha(\mathbf{x})$  shows that  $U_\alpha$  is concave for any  $\alpha \in (0, \infty)$ . Thus [17, p. 69],  $U_\alpha(\mathbf{x}_\alpha^*) \leq U_\alpha(\mathbf{x}) + \nabla U_\alpha(\mathbf{x})^T (\mathbf{x}_\alpha^* - \mathbf{x})$ , which yields

$$U_\alpha(\mathbf{x}_\alpha^*) < U_\alpha(\mathbf{x}). \quad (2.9)$$

This with (2.8) establish the desired contradiction.  $\square$

Proposition 1 shows that for an arbitrary set  $\mathcal{C}$  and  $M = 2$ , the  $\alpha$ -fair policy yields a tradeoff that is optimal from Jain's index perspective. However, this result does not necessarily carry over to cases with  $M > 2$  users. To show this, we constructed counter examples for  $M = 3$  and  $M = 4$ . The case of  $M = 4$  yields deeper insight and will be explained in more detail.

**Example 1.** Let  $\mathcal{C}$  contain two benefit vectors, i.e.,  $\mathcal{C} = \{\mathbf{x}, \mathbf{y}\}$ , where  $\mathbf{x} = [8, 8, 90, 90]$  and  $\mathbf{y} = [7, 14, 27, 86]$ .

For  $\alpha = 2$ , maximizing the  $\alpha$ -fair utility yields  $\mathbf{y}$  because  $U_2(\mathbf{y}) > U_2(\mathbf{x})$ . However,  $\eta(\mathbf{x}) = 196$ ,  $\eta(\mathbf{y}) = 134$ ,  $J(\mathbf{x}) = 0.59$  and  $J(\mathbf{y}) = 0.54$ , that is,  $\eta(\mathbf{x}) > \eta(\mathbf{y})$  and

$J(\mathbf{x}) > J(\mathbf{y})$ , which implies that  $\mathbf{x}$  is the optimal EJT benefit vector. This agrees with intuition since, by inspection,  $\mathbf{x}$  offers 75% of the users higher benefits than  $\mathbf{y}$ .  $\square$

Drawing more insight from the above example, we will show that the efficiency gap between the benefit vectors generated by the optimal EJT policy and those generated by the  $\alpha$ -fair one can be unbounded. To show that, let us modify  $\mathcal{C}$  in the above example such that  $\mathcal{C} = \{\mathbf{x}, \mathbf{y}, \hat{\mathbf{x}}, \hat{\mathbf{y}}\}$ , where  $\hat{\mathbf{x}} = c\mathbf{x}$ ,  $\hat{\mathbf{y}} = c\mathbf{y}$ , and  $c > 1$  is some constant. In this case, it can be easily verified that  $\hat{\mathbf{y}}$  is the  $\alpha$ -fair benefit vector and  $\hat{\mathbf{x}}$  is the optimal EJT benefit vector. Furthermore, because Jain's index is invariant under scaling,  $J(\hat{\mathbf{x}}) = J(\mathbf{x}) > J(\hat{\mathbf{y}}) = J(\mathbf{y})$ . However, direct computation reveals that  $\eta(\hat{\mathbf{x}}) - \eta(\hat{\mathbf{y}}) = c(\eta(\mathbf{x}) - \eta(\mathbf{y}))$ . Hence, an unbounded  $c$ , results in an unbounded difference in efficiency between the optimal EJT and the  $\alpha$ -fair benefit vectors. The existence of such  $c$  depends, of course, on  $\mathcal{C}$ . In fact, it will be shown later that the structure of  $\mathcal{C}$  is intimately related to the optimal EJT.

Another insight that can be drawn from the above example is that the  $\alpha$ -fair benefit vector corresponding to  $\alpha = 0$  is  $\mathbf{x}$ . From Jain's index perspective, this vector is fairer than the  $\alpha$ -fair benefit vector corresponding to  $\alpha = 2$ . Hence, this example shows that, although increasing  $\alpha$  results in benefit vectors that are fairer in the senses considered in [3] and [5], it does not necessarily improve fairness in the Jain's index sense.

Many applications, including wireless communications ones, involve the tradeoff between the benefit vectors of more than two users. Since in these cases, maximizing the  $\alpha$ -fair utility does not necessarily yield benefit vectors that achieve the optimal EJT (cf. Example 1), in the next section we will develop a technique for achieving this tradeoff.

## 2.4 The Optimal Efficiency-Jain Tradeoff Policy

In this section, we develop a generic technique for obtaining the optimal EJT for an arbitrary set  $\mathcal{C}$ . To enable practical implementation of this technique, we identify conditions on  $\mathcal{C}$ , which render the underlying optimization problems easy to solve. We will then provide instances in which these conditions are satisfied in practice. A geometric interpretation that commends the use of Jain's index as a fairness measure is then provided. We conclude this section by providing an alternate formulation that will prove useful in characterizing and achieving the optimal EJT in ergodic time-varying scenarios.

### 2.4.1 A Technique for Obtaining the Optimal EJT for an Arbitrary $\mathcal{C}$

A standard approach to find the optimum tradeoff benefit vectors is to use scalarization [17, p. 178]. Using scalarization, one can devise a parameterized utility function of the form

$$U_\nu(\mathbf{x}) = \nu\eta(\mathbf{x}) + (1 - \nu)F(\mathbf{x}), \quad (2.10)$$

where  $\nu \in [0, 1]$  is a parameter that controls the emphasis on efficiency relative to the emphasis on fairness, and  $F(\mathbf{x})$  is a fairness function that maps  $\mathbf{x}$  into a real number, to quantify fairness, such as Jain's index. If  $U_\nu(\mathbf{x})$  is a concave function, and  $\mathcal{C}$  is a convex set, then a benefit vector that maximizes  $U_\nu(\mathbf{x})$  achieves the optimum tradeoff between  $\eta(\mathbf{x})$  and  $F(\mathbf{x})$  and such a vector can be computed efficiently using standard convex optimization techniques. Unfortunately, it can be shown that for  $F(\mathbf{x}) = J(\mathbf{x})$ ,  $U_\nu(\mathbf{x})$  given by (2.10) is not concave. One might be tempted to replace Jain's index with another meaningful fairness function such that  $U_\nu(\mathbf{x})$  is concave. A meaningful fairness function should be homogeneous of degree zero [2,5], i.e., it should

be invariant under scaling so that it is not affected by changing the units under which the benefit vector is measured. Unfortunately, as we prove in the following Lemma, the only homogeneous function of degree zero that makes  $U_\nu(\mathbf{x})$  concave is a constant function, which is not a good fairness measure.

**Lemma 1.** *Let  $U_\nu(\mathbf{x}) : \mathbb{R}_+^M \rightarrow \mathbb{R}$  be a utility function that is parameterized by  $\nu \in [0, 1]$  such that  $U_\nu(\mathbf{x}) = \nu\eta(\mathbf{x}) + (1 - \nu)F(\mathbf{x})$ , where  $F(\mathbf{x})$  is a fairness function that is homogeneous of degree zero. Then,  $U_\nu(\mathbf{x})$  is a concave function if and only if  $F(\mathbf{x})$  is a constant function.*

*Proof.* We proceed by showing that if  $U_\nu(\mathbf{x})$  is concave, then  $F(\mathbf{x})$  is a constant function and showing the converse, i.e., if  $F(\mathbf{x})$  is a constant function, then  $U_\nu(\mathbf{x})$  is concave.

Assume that  $U_\nu(\mathbf{x})$  is a concave function. By this assumption, we have [17, p. 69]

$$U_\nu(\mathbf{y}) \leq U_\nu(\mathbf{x}) + \nabla U_\nu(\mathbf{x})^T (\mathbf{y} - \mathbf{x}), \quad \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}_+^M. \quad (2.11)$$

Since  $F(\mathbf{x})$  is a homogeneous function of degree zero, then

$$F(c\mathbf{y}) = F(\mathbf{y}), \quad \forall c > 0, \quad (2.12)$$

$$\nabla F(\mathbf{x})^T \mathbf{x} = 0, \quad (2.13)$$

where (2.12) follows from the definition of a homogeneous function of degree zero and (2.13) follows from Euler's homogeneous function theorem [19]. By substituting (2.12) and (2.13) in (2.11), we get

$$U_\nu(c\mathbf{y}) = c\nu\eta(\mathbf{y}) + (1 - \nu)F(\mathbf{y}) \leq U_\nu(\mathbf{x}) + c\nabla U_\nu(\mathbf{x})^T \mathbf{y} - \nu\mathbf{1}^T \mathbf{x}, \rightarrow \quad (2.14)$$

$$F(\mathbf{y}) \leq F(\mathbf{x}) + \frac{c}{(1-\nu)}(\nabla U_\nu(\mathbf{x}) - \nu\mathbf{1})^T \mathbf{y}, \quad \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}_+^M, \quad \forall c > 0.$$

We now show that  $F(\mathbf{x})$  is a constant function by contradiction. Suppose that  $F(\mathbf{x})$  is not a constant function, i.e., there exist  $\mathbf{x}$  and  $\mathbf{y}$  such that  $F(\mathbf{y}) > F(\mathbf{x}) + \epsilon, \epsilon > 0$ . We will show the contradiction by showing that there exists  $c$ , such that (2.14) does not hold. If  $(\nabla U_\nu(\mathbf{x}) - \nu \mathbf{1})^T \mathbf{y} < 0$ , then (2.14) does not hold when  $c > \frac{\epsilon(1-\nu)}{|(\nabla U_\nu(\mathbf{x}) - \nu \mathbf{1})^T \mathbf{y}|}$ . If  $(\nabla U_\nu(\mathbf{x}) - \nu \mathbf{1})^T \mathbf{y} > 0$ , then (2.14) does not hold when  $c < \frac{\epsilon(1-\nu)}{(\nabla U_\nu(\mathbf{x}) - \nu \mathbf{1})^T \mathbf{y}}$ . If  $(\nabla U_\nu(\mathbf{x}) - \nu \mathbf{1})^T \mathbf{y} = 0$ , then (2.14) does not hold for any  $c$ . Therefore, (2.14) holds true only when  $F(\mathbf{y}) = F(\mathbf{x}), \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}_+^M$ , i.e.,  $F(\mathbf{x})$  is a constant function.

Conversely, if  $F(\mathbf{x})$  is a constant function, then  $U_\nu(\mathbf{x})$  is affine and thus concave. □

**Corollary 1.** *The only fairness function,  $F(x)$ , that is homogeneous of degree zero and concave is the constant function.*

*Proof.* The proof follows directly from Lemma 1 by using the fact that  $F(x) = U_{\nu=0}(x)$ . □

Based on the result of Lemma 1, we conclude that it is not straightforward to use scalarization in our problem, even if  $\mathcal{C}$  is a convex set.

Another approach to find the optimal EJT is based on maximizing Jain's index while constraining the efficiency to be greater than a parameterized threshold. To elaborate, let  $\sigma$  be a threshold on the minimum efficiency, and let  $\mathcal{X}_\sigma$  be the set of all benefit vectors that yield an efficiency greater than  $\sigma$  and, at the same time, maximize Jain's index, that is,

$$\mathcal{X}_\sigma \triangleq \left\{ \mathbf{x} \mid \mathbf{x} = \arg \max_{\eta(\mathbf{x}) \geq \sigma, \mathbf{x} \in \mathcal{C}} J(\mathbf{x}) \right\}. \quad (2.15)$$

We note that the cardinality of  $\mathcal{X}_\sigma$  depends on  $\mathcal{C}$ . Furthermore, some elements in  $\mathcal{X}_\sigma$  may satisfy the condition  $\eta(\mathbf{x}) \geq \sigma$  in (2.15) with a strict inequality. Since we are seeking the benefit vectors that achieve the optimal EJT, we pick those vectors in  $\mathcal{X}_\sigma$

that maximize  $\eta(\mathbf{x})$ . In particular, let  $\mathbf{x}_\sigma^*$  be one of the benefit vectors that achieve the optimal EJT corresponding to  $\sigma$ , that is,

$$\mathbf{x}_\sigma^* \in \arg \max_{\mathbf{x} \in \mathcal{X}_\sigma} \eta(\mathbf{x}). \quad (2.16)$$

From (2.15) and (2.16), it can be seen that, for the given  $\sigma$ ,  $\mathbf{x}_\sigma^*$  achieves the optimal EJT in Definition 2. Hence, the set of all EJT-optimal benefit vectors can be obtained by decrementing  $\sigma$  from  $\sigma_{\max} = \max_{\mathbf{x} \in \mathcal{C}} \eta(\mathbf{x})$  to  $\sigma_{\min} = \min_{\mathbf{x} \in \mathcal{C}} \eta(\mathbf{x})$  in  $P + 1$  steps, each of size  $\delta$ . For each step  $p$ ,  $p = 0, \dots, P$ , the optimization problems in (2.15) and (2.16) corresponding to  $\sigma = \sigma_{\max} - p\delta$  are solved; a smaller  $\delta$  results in evaluating more points and therefore obtaining a smoother EJT curve. This policy is presented formally in Procedure 1.

---

**Procedure 1** Optimal EJT policy for arbitrary  $\mathcal{C}$

---

**Input:** Arbitrary set  $\mathcal{C}$  and step  $\delta > 0$

**Output:**  $\mathbf{x}_\sigma^*$

- 1: Initialize  $\sigma_{\min} = \min_{\mathbf{x} \in \mathcal{C}} \eta(\mathbf{x})$ ,  $\sigma_{\max} = \max_{\mathbf{x} \in \mathcal{C}} \eta(\mathbf{x})$  and  $P = \lfloor (\sigma_{\max} - \sigma_{\min}) / \delta \rfloor$ .
  - 2: **for**  $p = 0 : P$ , **do**
  - 3:    $\sigma = \sigma_{\max} - p\delta$
  - 4:   Find  $\mathcal{X}_\sigma$  in (2.15).
  - 5:    $\mathbf{x}_\sigma^* \leftarrow \arg \max_{\mathbf{x} \in \mathcal{X}_\sigma} \eta(\mathbf{x})$ .
  - 6: **end for**
- 

Inspection of Procedure 1 reveals that the main difficulty in obtaining  $\mathbf{x}_\sigma^*$  lies in finding a solution of the optimization problem in (2.15), let alone finding the entire set  $\mathcal{X}_\sigma$ . This difficulty arises because  $J(\mathbf{x})$  is a non-concave function, even when  $\mathcal{C}$  is a convex set<sup>2</sup>. Indeed, any non-constant homogeneous fairness function with degree zero is a non-concave function as given by Corollary 1. If the dimension of the set of feasible benefits is large, Procedure 1 can be prohibitively complex to

---

<sup>2</sup>It can be shown that  $J(\mathbf{x})$  is a (non-strictly) quasi-concave function; thus, using bisection and convex feasibility optimization techniques, it is possible to find efficiently an optimal solution when  $\mathcal{C}$  is convex [17, pp. 145–146]. However, it is still difficult to find the entire set of optimal solutions  $\mathcal{X}_\sigma$ , for quasi-concave objective functions even when  $\mathcal{C}$  is convex.

implement in real-time scenarios. In such cases, this procedure might be used as a benchmark for less costly algorithms that approximate the solution of the underlying non-convex optimization problems. The accuracy of such algorithms depends on the approximation technique and the properties of  $\mathcal{C}$ . The complexity of Procedure 1 motivates us to seek conditions on  $\mathcal{C}$  that enable the optimal EJT to be readily obtained.

We conclude this section by highlighting that finding the optimal EJT benefit vectors can also be formulated based on finding all benefit vectors that maximize the efficiency while constraining the Jain's index to be greater than a parameterized threshold, and then choose those vectors that have the highest Jain's index. That is, for a given threshold on the minimum Jain's index,  $J_0$ , the benefit vectors that achieve the optimal EJT,  $\mathbf{x}_{J_0}^*$ , can be found as follows

$$\mathbf{x}_{J_0}^* \in \arg \max_{\mathbf{x} \in \mathcal{X}_{J_0}} J(\mathbf{x}), \quad (2.17)$$

where

$$\mathcal{X}_{J_0} \triangleq \left\{ \mathbf{x} \mid \mathbf{x} = \arg \max_{J(\mathbf{x}) \geq J_0, \mathbf{x} \in \mathcal{C}} \eta(\mathbf{x}) \right\}. \quad (2.18)$$

If  $\mathcal{C}$  is convex, then finding a single element of  $\mathcal{X}_{J_0}$  can be cast as a second-order cone programming problem, which can be solved efficiently using standard convex optimization techniques [20]. However, such an approach has two main disadvantages. First of all, finding all elements of the set  $\mathcal{X}_{J_0}$  is still difficult even if  $\mathcal{C}$  is convex<sup>3</sup>. Secondly, and most importantly, it is shown in [20] that it is difficult to incorporate such formulation in the problem of scheduling radio resources in time-varying channels. Indeed, it is demonstrated that constructing a scheduler based on this formulation leads to inferior EJT as compared to  $\alpha$ -fair based scheduling, cf. [20, Fig. 2]. Since

---

<sup>3</sup>Note that the objective function in (2.18) is not strictly concave which implies that the solution for (2.18) is not necessarily unique.

the problem of scheduling radio resources in time-varying channels is the focus of this thesis, the formulation given by (2.17) and (2.18) is not considered hereinafter.

### 2.4.2 A Property for Ensuring Tractability

In order to render the optimization problems underlying (2.15) easy to solve, we begin by identifying a class of sets  $\mathcal{C}$  which satisfy what we refer to as the “monotonic tradeoff property”. To do so, let  $J_\sigma^*$  denote the maximum Jain’s index corresponding to an efficiency  $\eta(\mathbf{x}) = \sigma$ , i.e.,

$$J_\sigma^* = \max_{\eta(\mathbf{x})=\sigma, \mathbf{x} \in \mathcal{C}} J(\mathbf{x}). \quad (2.19)$$

By definition,  $J_\sigma^*$  is unique. However, it might be achieved by multiple benefit vectors.

Using (2.19), we are now ready to define the monotonic tradeoff property.

**Definition 3** (Monotonic Tradeoff Property). *A set  $\mathcal{C}$  is said to possess the monotonic tradeoff property if  $J_\sigma^*$  is strictly decreasing in  $\sigma$ , for  $\sigma \geq \sigma^*$ , and constant otherwise.*

□

This definition states that a set that possesses the monotonic tradeoff property is one in which any decrease in efficiency results in a strict increase in Jain’s index, until  $\sigma^*$  is reached. Decreasing efficiency beyond  $\sigma^*$  maintains Jain’s index at its maximum. In other words,

$$J_{\sigma^*}^* = \max_{\eta(\mathbf{x})=\sigma^*, \mathbf{x} \in \mathcal{C}} J(\mathbf{x}) = \max_{\mathbf{x} \in \mathcal{C}} J(\mathbf{x}). \quad (2.20)$$

An instance in which  $\mathcal{C}$  satisfies the monotonic tradeoff property is shown in Fig. 2.2(a) and the corresponding EJT is shown in Fig. 2.2(b). These figures will be discussed in the next section.

We will now show how the monotonic tradeoff property facilitates finding the benefit vectors that achieve the optimal EJT. When a set possesses this property and  $\sigma > \sigma^*$ , the inequality  $\eta(\mathbf{x}) \geq \sigma$  in (2.15) is satisfied with equality because  $J_\sigma^*$  is strictly decreasing in  $\sigma$ . In this case, the optimization in (2.15) is equivalent to that in (2.19). We now use (2.19) to obtain an equivalent convex formulation. By definition,  $J(\mathbf{x}) = \frac{\eta^2(\mathbf{x})}{M\|\mathbf{x}\|^2}$ . Hence, when  $\eta(\mathbf{x}) = \sigma$ , the objective in (2.19) can be expressed as  $\frac{\sigma^2}{M\|\mathbf{x}\|^2}$  and (2.19) can be cast in the following equivalent form:

$$\min_{\eta(\mathbf{x})=\sigma, \mathbf{x} \in \mathcal{C}} \|\mathbf{x}\|^2. \quad (2.21)$$

In contrast with (2.15), the objective in (2.21) is convex. In fact, this objective is strictly convex, which implies that when  $\mathcal{C}$  too is convex, the optimization problem in (2.21) is easy to solve and its solution is unique [17, p. 397]. In addition, if  $\mathcal{C}$  is not convex and (2.21) has multiple solutions, all these solutions will achieve the same EJT as they all have the same efficiency,  $\sigma$ , and the same Jain's index. This observation eliminates the requirement for finding all solutions in (2.15) since any solution of (2.21) achieves the optimal EJT. To summarize, if the monotonic tradeoff property in Definition 3 is satisfied,  $\mathbf{x}_\sigma^*$  can be found by solving (2.21), which is significantly easier to solve than the optimization problems in (2.15) and (2.16) for an arbitrary  $\mathcal{C}$ .

Similar to Procedure 1, the benefit vectors that achieve the optimal EJT can be obtained by varying  $\sigma$  from  $\sigma_{\max}$  to  $\sigma_{\min}$ . However, when  $\mathcal{C}$  possesses the monotonic tradeoff property, we can find  $\mathbf{x}_\sigma^*$  by solving (2.21) for each  $\sigma$ . This policy is presented formally in Procedure 2 below.

---

**Procedure 2** Optimal EJT for  $\mathcal{C}$  possessing the monotonic tradeoff property

---

**Input:** A set  $\mathcal{C}$  possessing the monotonic tradeoff property and step  $\delta > 0$

**Output:**  $\mathbf{x}_\sigma^*$

- 1: Initialize  $\sigma_{\min} = \min_{\mathbf{x} \in \mathcal{C}} \eta(\mathbf{x})$ ,  $\sigma_{\max} = \max_{\mathbf{x} \in \mathcal{C}} \eta(\mathbf{x})$  and  $P = \lfloor (\sigma_{\max} - \sigma_{\min}) / \delta \rfloor$ .
  - 2: **for**  $p = 0 : P$ , **do**
  - 3:    $\sigma = \sigma_{\max} - p\delta$
  - 4:    $\mathbf{x}_\sigma^* = \arg \min_{\eta(\mathbf{x})=\sigma, \mathbf{x} \in \mathcal{C}} \|\mathbf{x}\|^2$
  - 5:   **if**  $J(\mathbf{x}_\sigma^*) = J(\mathbf{x}_{\sigma+\delta}^*)$  **then**
  - 6:     quit
  - 7:   **end if**
  - 8: **end for**
- 

### 2.4.3 Sufficient Conditions for Satisfying the Monotonic Tradeoff Property

In the previous section, we showed that finding the set of benefit vectors that achieve the optimal EJT is significantly simplified when the set  $\mathcal{C}$  possesses the monotonic tradeoff property. Unfortunately, we have not been able to identify a distinguishing feature that is necessary for a set to possess that property. For instance, the monotonic tradeoff property can be possessed by sets that are either continuous or discrete, convex or otherwise. This observation suggests that deriving necessary conditions might be elusive. However, we have been able to obtain sufficient conditions that ensure that a given set possesses this property. Such conditions are provided in Theorem 1 below<sup>4</sup>.

**Theorem 1.** *The set  $\mathcal{C}$  possesses the monotonic tradeoff property if:*

- i.  $\mathcal{C}$  is convex;*
- ii.  $x_{\min} \mathbf{1}_M \in \mathcal{C}$ ; and*
- iii. every  $\mathbf{x} \in \mathcal{C}$  satisfies  $\mathbf{x} \succeq x_{\min} \mathbf{1}_M$ , where  $x_{\min} \geq 0$  provides a guarantee on the*

---

<sup>4</sup>This theorem is a generalized version of the one we provided in [21], wherein  $x_{\min}$  was restricted to be zero.

minimum benefit that each user receives.

*Proof.* Let  $\mathbf{x}_{\sigma_1}^*$  and  $\mathbf{x}_{\sigma_2}^*$  be the benefit vectors obtained using (2.21) with  $\sigma_1$  and  $\sigma_2$ , respectively, where  $\sigma_{\min} \leq \sigma_1 < \sigma_2 \leq \sigma_{\max}$ . To prove Theorem 1, it suffices to show that if the conditions of the theorem are satisfied, then  $J_{\sigma_1}^* \geq J_{\sigma_2}^*$  with equality if and only if  $J_{\sigma_2}^* = J_{\sigma^*}^*$ , where  $J_{\sigma}^*$  and  $J_{\sigma^*}^*$  are given by (2.19) and (2.20), respectively. We consider two distinct cases:  $x_{\min} > 0$  and  $x_{\min} = 0$ .

**Case 1** ( $x_{\min} > 0$ ) First we note that, because  $x_{\min}\mathbf{1} \in \mathcal{C}$ ,  $J_{\sigma^*}^* = 1$ . Now, the convexity of  $\mathcal{C}$  implies that, for any  $\theta \in (0, 1)$ ,

$$\mathbf{y}_\theta = \theta x_{\min}\mathbf{1} + (1 - \theta)\mathbf{x}_{\sigma_2}^* \in \mathcal{C}. \quad (2.22)$$

Substituting for  $\mathbf{y}_\theta$  from (2.22) in (2.1), it can be verified that

$$J(\mathbf{y}_\theta) - J_{\sigma_2}^* = a(1 - J(\mathbf{y}_\theta)), \quad (2.23)$$

where  $a = \frac{\theta^2 M^2 x_{\min}^2 + 2\theta(1-\theta)Mx_{\min}\mathbf{1}^T \mathbf{x}_{\sigma_2}^*}{M(1-\theta)^2 \mathbf{x}_{\sigma_2}^{*T} \mathbf{x}_{\sigma_2}^*} \geq 0$ .

To prove the theorem, we will show that  $J(\mathbf{y}_\theta) - J_{\sigma_2}^* \geq 0$ , with equality if and only if  $J_{\sigma_2}^* = 1$ . To do so, we note that, by the properties of Jain's index, if  $J_{\sigma_2}^* < 1$ , then  $\mathbf{x}_{\sigma_2}^* \neq \gamma_1 \mathbf{1}$ , and subsequently from (2.22)  $\mathbf{y}_\theta \neq \gamma_2 \mathbf{1}$  for any  $\gamma_1, \gamma_2 > 0$ . This implies that  $J(\mathbf{y}_\theta) < 1$ . This with (2.23) and the fact that  $a > 0$  for any  $\theta \in (0, 1)$  further implies that  $J(\mathbf{y}_\theta) > J_{\sigma_2}^*$ . We next show that there exists a  $\theta_0 \in (0, 1)$  such that  $\eta(\mathbf{y}_{\theta_0}) = \sigma_1$ . In particular, setting  $\theta_0 = \frac{\sigma_2 - \sigma_1}{\sigma_2 - Mx_{\min}}$  yields this efficiency. Since  $J_{\sigma_1}^*$  is the maximum Jain's index corresponding to  $\eta(\mathbf{x}) = \sigma_1$ , we must have  $J_{\sigma_1}^* \geq J(\mathbf{y}_{\theta_0})$ , which implies that  $J_{\sigma_1}^* > J_{\sigma_2}^*$ .

For the case of  $J_{\sigma_2}^* = 1$ , we note that  $\mathbf{y}_\theta = \gamma_2 \mathbf{1}$ , for some  $\gamma_2$  and therefore, by the homogeneity of Jain's index,  $J(\mathbf{y}_\theta) = 1$  for any  $\theta$ . The statement of the theorem follows by using an argument analogous to the above and noting that  $J_{\sigma}^* \leq 1$ , for any

$\sigma$  including  $\sigma_1$ .

**Case 2** ( $x_{\min} = 0$ ) The proof for the case of  $x_{\min} > 0$  does not hold when  $x_{\min} = 0$ . Furthermore, it is easy to find sets  $\mathcal{C}$  for which  $x_{\min} = 0$ , but vectors of the form  $\gamma_1 \mathbf{1} \notin \mathcal{C}$  even for infinitesimal values of  $\gamma_1 > 0$ , e.g.,  $\mathcal{C} = \{x_1, x_2 : 0 \leq x_2 \leq cx_1\}$ , where  $c \in (0, 1)$ .

To prove the theorem for this case, we provide the following lemma:

**Lemma 2.** *Let  $\mathcal{C}$  be a set satisfying the conditions of Theorem 1 with  $x_{\min} = 0$  and let  $\sigma_{\min} \leq \sigma_1 \leq \sigma_2 \leq \sigma_{\max}$ , then  $J_{\sigma_1}^* \geq J_{\sigma_2}^*$ .*

Before providing the proof, we note that unlike the claim of Theorem 1, this lemma does not tell whether a strict decrease in  $\sigma$  will yield a strict increase in Jain's index.

*Proof.* Let  $\mathbf{x}_{\sigma_2}^*$  be the optimal benefit vector at efficiency  $\sigma_2$ ; that is,  $J(\mathbf{x}_{\sigma_2}^*) = J_{\sigma_2}^*$ . Since  $\mathcal{C}$  is convex and  $\mathbf{0}_M \in \mathcal{C}$ , it follows that, for any  $\theta \in [0, 1]$ ,  $\hat{\mathbf{x}} = \theta \mathbf{x}_{\sigma_2}^* \in \mathcal{C}$ . Noting that  $\sigma_1 \leq \sigma_2$  and setting  $\theta = \frac{\sigma_1}{\sigma_2}$ , it can be seen that  $\eta(\hat{\mathbf{x}}) = \sigma_1$ . Now, the homogeneity of Jain's index implies that  $J(\hat{\mathbf{x}}) = J_{\sigma_2}^*$ . However,  $J_{\sigma_1}^*$  is the optimal Jain's index at efficiency  $\sigma_1$  and hence  $J_{\sigma_1}^* \geq J(\hat{\mathbf{x}})$ , which completes the proof of the lemma.  $\square$

We now proceed to complete the proof of Theorem 1. Let  $\sigma^*$  be the highest efficiency corresponding to the maximum achievable Jain's index, i.e.,  $J_{\sigma^*}^*$ . We will consider two cases:  $\sigma_2 \leq \sigma^*$  and  $\sigma_2 > \sigma^*$ .

When  $\sigma_2 \leq \sigma^*$ , we have by assumption that  $\sigma_1 < \sigma^*$ , which, by Lemma 2, implies that  $J_{\sigma_2}^* = J_{\sigma_1}^* = J_{\sigma_2}^*$ , and the theorem is proved in this case.

We now consider the case of  $\sigma_2 > \sigma^*$ . We again consider two cases:  $\sigma_1 \leq \sigma^*$  and  $\sigma_1 > \sigma^*$ .

Since  $\sigma_2 > \sigma^*$ , it follows by the definition of  $\sigma^*$  that  $J_{\sigma_2}^* < J_{\sigma^*}^*$ . Now, if  $\sigma_1 \leq \sigma^*$ , invoking Lemma 2 yields  $J_{\sigma_1}^* = J_{\sigma^*}^*$  which implies that  $J_{\sigma_2}^* < J_{\sigma_1}^*$ .

We next prove the theorem for the remaining case of  $\sigma_2 > \sigma_1 > \sigma^*$ .

For this case, we will consider the line segment,  $\mathbf{y}_\theta$ , connecting  $\mathbf{x}_{\sigma_2}^*$  with  $\mathbf{x}_{\sigma^*}^*$ , i.e.,  $\mathbf{y}_\theta = \theta \mathbf{x}_{\sigma^*}^* + (1 - \theta) \mathbf{x}_{\sigma_2}^*$ ,  $\theta \in [0, 1]$ . The convexity assumption implies that  $\mathbf{y}_\theta \in \mathcal{C}$ . Substituting for  $\mathbf{y}_\theta$  in (2.1), it can be verified that, for any  $\mathbf{x}_{\sigma_2}^*$  and  $\mathbf{y}_\theta$  such that  $\|\mathbf{x}_{\sigma_2}^*\| \|\mathbf{y}_\theta\| > 0$ ,

$$J(\mathbf{y}_\theta) - J(\mathbf{x}_{\sigma_2}^*) = \frac{\theta^2(\sigma^{*2}\|\mathbf{x}_{\sigma_2}^*\|^2 - \sigma_2^2\|\mathbf{x}_{\sigma^*}^*\|^2) + 2\theta(1 - \theta)\sigma_2(\sigma^*\|\mathbf{x}_{\sigma_2}^*\|^2 - \sigma_2\mathbf{x}_{\sigma^*}^{*T}\mathbf{x}_{\sigma_2}^*)}{M\|\mathbf{x}_{\sigma_2}^*\|^2\|\mathbf{y}_\theta\|^2}. \quad (2.24)$$

We will now use this equality to show that  $J(\mathbf{y}_\theta) > J(\mathbf{x}_{\sigma_2}^*)$ . For the first term in the numerator, we note that, since by assumption  $\sigma_2 > \sigma^*$ , we have  $J_{\sigma_2}^* < J_{\sigma^*}^*$ . This implies that  $\sigma^{*2}\|\mathbf{x}_{\sigma_2}^*\|^2 - \sigma_2^2\|\mathbf{x}_{\sigma^*}^*\|^2 > 0$ . For the second term, we note that, by the Cauchy-Schwartz inequality,

$$\sigma^*\|\mathbf{x}_{\sigma_2}^*\|^2 - \sigma_2\mathbf{x}_{\sigma^*}^{*T}\mathbf{x}_{\sigma_2}^* \geq \|\mathbf{x}_{\sigma_2}^*\|(\sigma^*\|\mathbf{x}_{\sigma_2}^*\| - \sigma_2\|\mathbf{x}_{\sigma^*}^*\|) > 0.$$

The strict positivity follows from the fact that  $J_{\sigma_2}^* < J_{\sigma^*}^*$ . Hence, we have shown that  $J(\mathbf{y}_\theta) > J(\mathbf{x}_{\sigma_2}^*)$  for any  $\theta \in (0, 1)$ .

We next show that there exists a  $\theta_0 \in (0, 1)$  such that  $\eta(\mathbf{y}_{\theta_0}) = \sigma_1$ . In particular, setting  $\theta_0 = \frac{\sigma_2 - \sigma_1}{\sigma_2 - \sigma^*}$  yields this efficiency, where  $\theta_0 \in (0, 1)$  by the assumption that  $\sigma_1 > \sigma^*$ . Since  $J_{\sigma_1}^*$  is the maximum Jain's index corresponding to  $\eta(\mathbf{x}) = \sigma_1$ , we must have  $J_{\sigma_1}^* \geq J(\mathbf{y}_{\theta_0})$ , which implies that  $J_{\sigma_1}^* > J_{\sigma_2}^*$  and completes the proof of the theorem.  $\square$

To provide a graphical illustration of Theorem 1, in Fig. 2.2(a) we show a feasible set  $\mathcal{C}$  satisfying the conditions of the theorem with  $x_{\min} = 0$  for a case with  $M = 2$  users. The EJT corresponding to the set in Fig. 2.2(a) is shown in Fig. 2.2(b).

To show how Fig. 2.2(b) is obtained, we begin by noting that, in Fig 2.2(a), the maximum Jain's fairness line  $x_1 = x_2$  passes through  $\mathcal{C}$  and yields  $J(\mathbf{x}) = 1$ . The regular-weight dashed lines in this figure represent the constant efficiency levels,

$\eta(\mathbf{x}) = \sigma$ , at different values of  $\sigma$ . For  $\sigma \leq 5.33$ , the points at which the dashed lines intersect the  $x_1 = x_2$  line lie inside  $\mathcal{C}$ . In this case, the maximal Jain's index,  $J_\sigma^* = 1$ . For  $\sigma > 5.33$ , the dashed lines representing the  $\eta(\mathbf{x}) = \sigma$  levels intersect the  $x_1 = x_2$  line at points outside  $\mathcal{C}$ . For these efficiency levels, the maximal Jain's indices are strictly less than 1 and correspond to the points at which the dashed lines intersect with the boundary of  $\mathcal{C}$ . The set of optimal EJT benefit vectors is shown by the thick dashed line on the boundary of  $\mathcal{C}$ . The variation of  $J_\sigma^*$  with  $\sigma$  is depicted in Fig. 2.2(b).

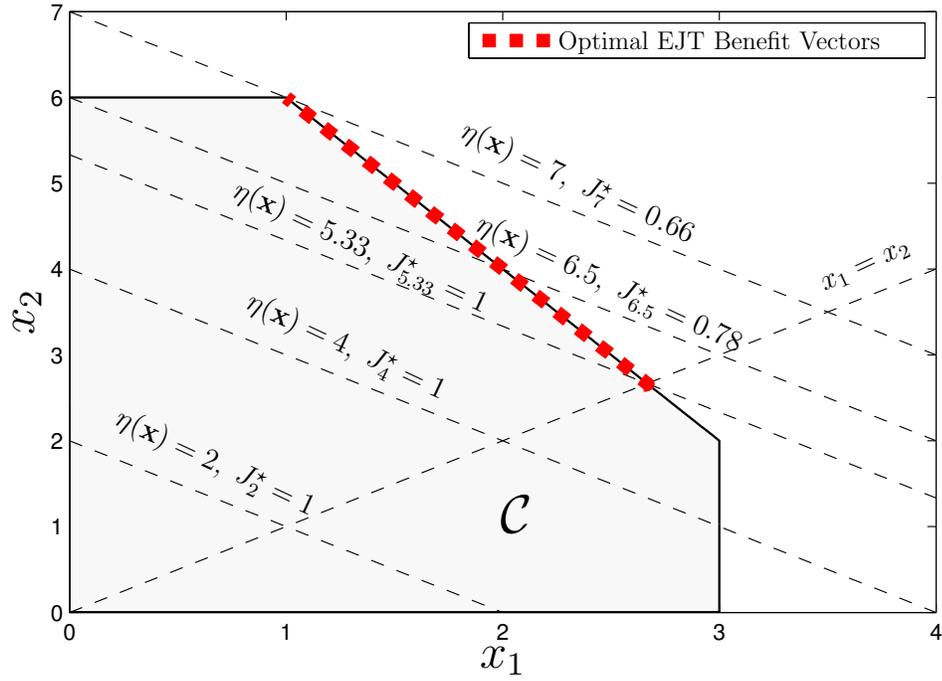
It can be seen from this figure that, in agreement with Theorem 1, the set  $\mathcal{C}$  satisfies the monotonic tradeoff property in Definition 3 with  $\sigma^* = 5.33$ . In this figure, the optimal EJT corresponding to the thick dashed line on the boundary of  $\mathcal{C}$  in Fig. 2.2(a) is represented by the thick dashed line to the right of  $\sigma^*$ .

Although necessary conditions are not available, the sufficient conditions provided in Theorem 1 are relatively sharp. To illustrate that, we consider the optimal EJT for the set  $\mathcal{C}$  shown in Fig. 2.3(a). This set satisfies the first condition of Theorem 1, but does not satisfy the second and third conditions. In other words,  $\mathcal{C}$  is convex, but there is no  $x_{\min}$  such that  $x_{\min} \mathbf{1}_M \in \mathcal{C}$  and  $x \succeq x_{\min} \mathbf{1}_M, \forall \mathbf{x} \in \mathcal{C}$ . We will now demonstrate that this set does not possess the monotonic tradeoff property in Definition 3.

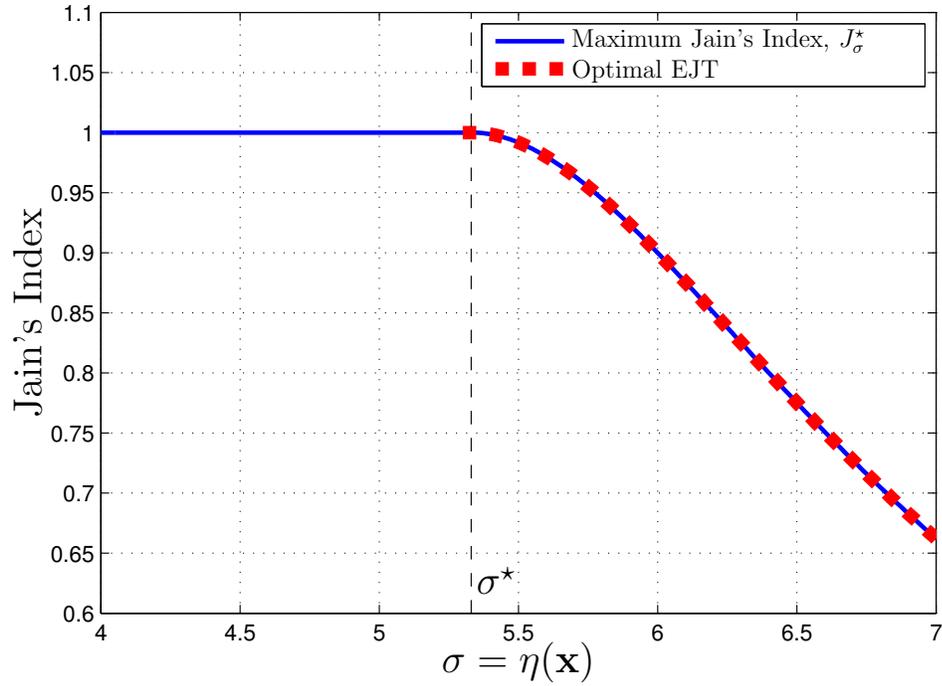
For the set shown in Fig. 2.3(a), the maximum Jain's fairness line  $x_1 = x_2$  intersects  $\mathcal{C}$  at one point, viz.,  $x_1 = x_2 = 6$ . At this point, the efficiency,  $\sigma = 12$  and Jain's index,  $J(\mathbf{x}) = 1$ . At any other point in  $\mathcal{C}$ , Jain's index is strictly less than 1. To see why this implies that  $\mathcal{C}$  does not possess the monotonic tradeoff property, we note that, for each dashed line representing constant  $\sigma \in [3.6, 12)$  and  $\sigma \in (12, 16]$  the maximal Jain's index,  $J_\sigma^*$ , corresponds to the intersection of the dashed line with the non-vertical part of the boundary of  $\mathcal{C}$ . For  $\sigma \in [3.6, 12)$ ,  $J_\sigma^*$  is strictly monotonically increasing in  $\sigma$ , and for  $\sigma \in (12, 16]$ ,  $J_\sigma^*$  is strictly monotonically decreasing in  $\sigma$ .<sup>5</sup>

---

<sup>5</sup>For this  $\mathcal{C}$ ,  $\sigma < 3.6$  and  $\sigma > 16$  are not feasible.



(a) A convex set that satisfies the conditions in Theorem 1.



(b) Variations of  $J_\sigma^*$  with  $\sigma$ .

**Figure 2.2:** The monotonic tradeoff property: An illustrative example.

Hence, it can be seen that, for  $\sigma < 12$ , the tradeoff is not meaningful, since, in that region, both  $\eta(\mathbf{x})$  and  $J_\sigma^*(\mathbf{x})$  can be increased at the same time. The optimal EJT benefit vectors are shown by the thick dashed line on the boundary of  $\mathcal{C}$ .

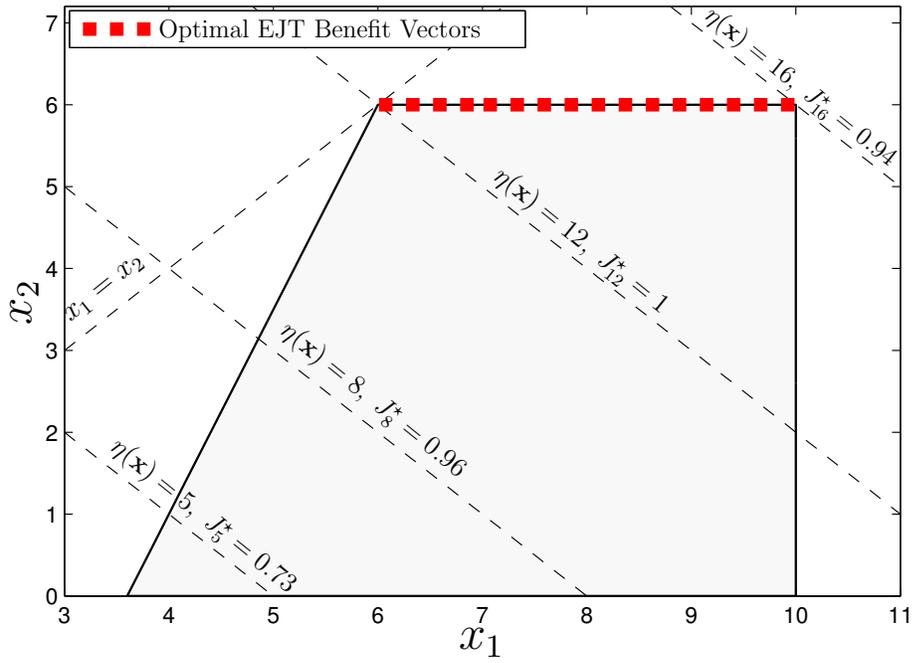
The variation of  $J_\sigma^*$  with  $\sigma$  is depicted in Fig. 2.3(b). As we pointed out,  $J_\sigma^*$  is strictly increasing for  $\sigma < 12$ , implying that  $\mathcal{C}$  does not satisfy the monotonic tradeoff property in Definition 3. For  $\sigma \in (12, 16]$ , the optimal EJT corresponding to the thick dashed line on the boundary of  $\mathcal{C}$  in Fig. 2.3(a) is represented by the thick dashed line to the right of  $\sigma^*$ .

#### 2.4.4 Geometric Interpretation of the Optimal EJT

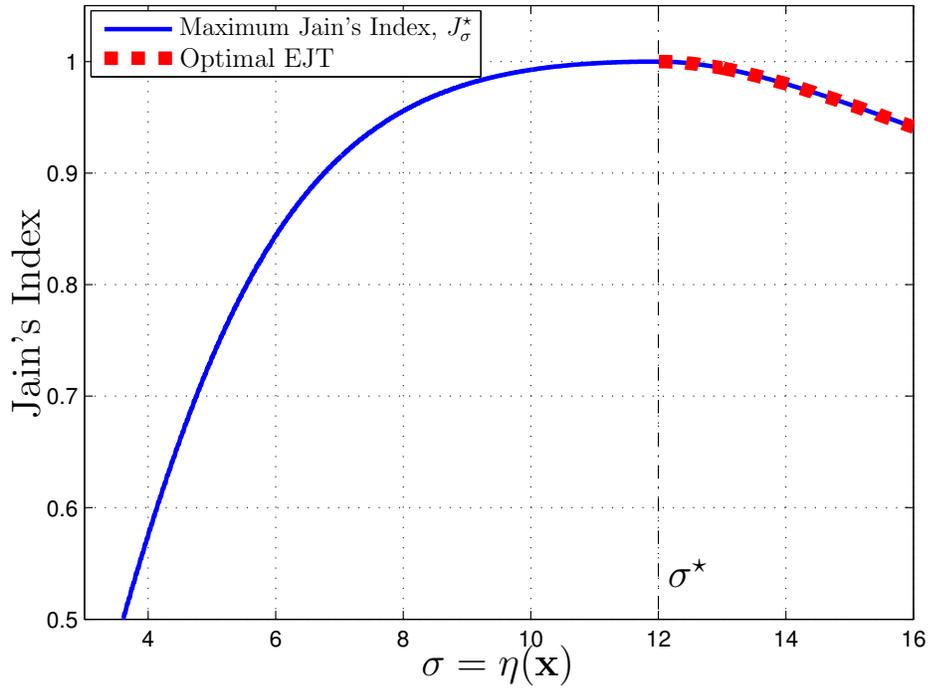
When  $\mathcal{C}$  satisfies the sufficient conditions given in Theorem 1, optimal EJT benefit vectors  $\{\mathbf{x}_\sigma^*\}$  have an interesting geometric interpretation. To see that, we use (2.21) to write

$$\begin{aligned} \mathbf{x}_\sigma^* &= \arg \min_{\eta(\mathbf{x})=\sigma, \mathbf{x} \in \mathcal{C}} \sum_{m=1}^M x_m^2 \\ &= \arg \min_{\eta(\mathbf{x})=\sigma, \mathbf{x} \in \mathcal{C}} \sum_{m=1}^M \left( x_m^2 - 2\frac{\sigma}{M}\eta(\mathbf{x}) + \frac{\sigma^2}{M^2} \right) \\ &= \arg \min_{\eta(\mathbf{x})=\sigma, \mathbf{x} \in \mathcal{C}} \left\| \mathbf{x} - \frac{\sigma}{M} \mathbf{1}_M \right\|^2. \end{aligned} \tag{2.25}$$

The last equality states that  $\mathbf{x}_\sigma^*$  is the unique Euclidean projection [17, p. 397] of the equal allocation vector  $\frac{\sigma}{M} \mathbf{1}_M$  onto the set  $\{\mathbf{x} | \eta(\mathbf{x}) = \sigma, \mathbf{x} \in \mathcal{C}\}$ . In other words, a benefit vector  $\mathbf{x}_\sigma^*$  achieves the optimal EJT if there is no feasible benefit vector  $\mathbf{y} \neq \mathbf{x}_\sigma^*$  such that  $\eta(\mathbf{y}) = \sigma$  is closer to the fairest solution  $\frac{\sigma}{M} \mathbf{1}_M$ . This interpretation commends the use of Jain's index as a fairness measure and is illustrated in Fig. 2.4. It also complements the interpretation given in [22] that Jain's index represents the angular deviation from a scaled all-one vector.

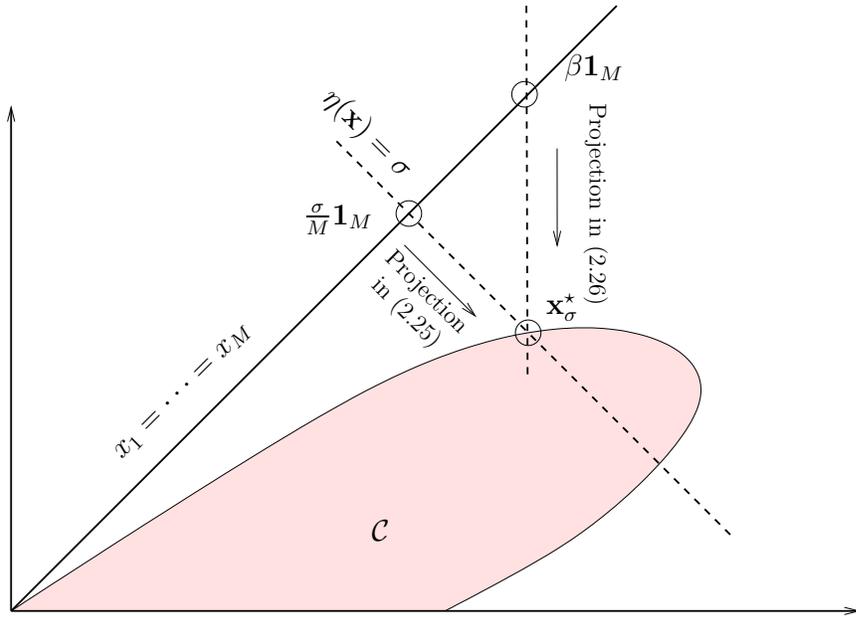


(a) A convex set that does not satisfy the conditions in Theorem 1.



(b) Variations of  $J_\sigma^*$  with  $\sigma$ .

**Figure 2.3:** An example of a set that does not possess the monotonic tradeoff property.



**Figure 2.4:** The optimal EJT benefit vector,  $\mathbf{x}_\sigma^*$ , is the unique projection of the fairest vector  $\frac{\sigma}{M}\mathbf{1}_M$  onto the set  $\{\mathbf{x}|\eta(\mathbf{x}) = \sigma, \mathbf{x} \in \mathcal{C}\}$ . Projection of  $\beta\mathbf{1}_M$  onto  $\mathcal{C}$  is also shown.

### 2.4.5 An Alternate Formulation

In Section 2.4.2 it was shown that, if the set  $\mathcal{C}$  possesses the monotonic tradeoff property, the optimal EJT benefit vectors can be obtained by solving the optimization problem in (2.21) for each  $\sigma$ . As such, the solution of (2.21) can be viewed as being parameterized by  $\sigma$ .

Although the form in (2.21) is convenient for providing an explicit characterization of the optimal EJT, the equality constraint therein renders it difficult to utilize in some applications. An instance of these applications is considered in the next section, wherein the instantaneous allocations of radio resources are to be updated for optimizing long-term average rates.

One approach to address the aforementioned difficulty is to incorporate the equality constraint into the objective by eliminating one of the variables [17, pp. 523–524]. However, this approach can be shown to result in complicating the computation of

gradient vectors necessary for the development of effective procedures for updating resource allocations. To alleviate this difficulty, we now provide an alternate formulation for (2.21). In this formulation, the efficiency  $\sigma$  is implicitly accounted for by a non-negative parameter  $\beta$  in the objective, and the benefit vectors are only constrained to lie in  $\mathcal{C}$ . In particular, when the conditions of Theorem 1 are satisfied, we have that for any  $\sigma = \eta(\mathbf{x})$ , the formulation in (2.21) is equivalent to

$$\min_{\mathbf{x} \in \mathcal{C}} \|\mathbf{x}\|^2 - 2\beta\eta(\mathbf{x}), \quad (2.26)$$

for some  $\beta \in [0, \infty)$ . To see this, we let  $\sigma$  be the efficiency corresponding to the solution of (2.26) for a given  $\beta \in [0, \infty)$ ; letting  $\beta = 0$  corresponds to  $\sigma_{\min}$ , the minimum feasible efficiency, and letting  $\beta \rightarrow \infty$  corresponds to  $\sigma \rightarrow \sigma_{\max}$ , the maximum feasible efficiency. To show the equivalence for intermediate values of  $\beta$ , we first show that the efficiency generated by (2.26) is strictly increasing in  $\beta$  in the following Lemma.

**Lemma 3.** *Let  $\mathbf{x}_\beta^* = \arg \min_{\mathbf{x} \in \mathcal{C}} \|\mathbf{x}\|^2 - 2\beta\eta(\mathbf{x})$ , where the set  $\mathcal{C}$  satisfies the conditions of Theorem 1 and  $\beta \in [0, \infty)$ . Then,  $\eta(\mathbf{x}_\beta^*)$  is strictly increasing in  $\beta$ .*

*Proof.* We will proceed by showing that for all  $\beta_1, \beta_2 \in [0, \infty)$ , such that  $\beta_1 > \beta_2$ , we have  $\eta(\mathbf{x}_{\beta_1}^*) > \eta(\mathbf{x}_{\beta_2}^*)$ . From the definition of  $\mathbf{x}_{\beta_1}^*$ ,  $\mathbf{x}_{\beta_2}^*$ , we can write the following strict inequalities:

$$\|\mathbf{x}_{\beta_1}^*\|^2 - 2\beta_1\eta(\mathbf{x}_{\beta_1}^*) < \|\mathbf{x}_{\beta_2}^*\|^2 - 2\beta_1\eta(\mathbf{x}_{\beta_2}^*), \quad (2.27)$$

$$\|\mathbf{x}_{\beta_2}^*\|^2 - 2\beta_2\eta(\mathbf{x}_{\beta_2}^*) < \|\mathbf{x}_{\beta_1}^*\|^2 - 2\beta_2\eta(\mathbf{x}_{\beta_1}^*). \quad (2.28)$$

The strict inequalities (2.27) and (2.28) follow from the unique optimality of  $\mathbf{x}_{\beta_1}^*$  and  $\mathbf{x}_{\beta_2}^*$ , respectively, where uniqueness is due to the strict convexity of the objective function and convexity of  $\mathcal{C}$  as per the conditions of Theorem 1. By adding (2.27)

and (2.28), we get following strict inequalities:

$$2(\beta_2 - \beta_1)\eta(\mathbf{x}_{\beta_1}^*) < 2(\beta_2 - \beta_1)\eta(\mathbf{x}_{\beta_2}^*) \Rightarrow \eta(\mathbf{x}_{\beta_1}^*) > \eta(\mathbf{x}_{\beta_2}^*), \quad (2.29)$$

where the last inequality follows since  $\beta_1 > \beta_2$ . □

Since  $\eta(\mathbf{x}_\beta^*)$  is strictly increasing in  $\beta$  according to Lemma 3, then the EJT obtained by letting  $\beta$  span the interval  $[0, \infty)$  in (2.26) is the same as that obtained by letting  $\sigma$  span the interval  $[\sigma_{\min}, \sigma_{\max})$  in (2.21).

Similar to the observation made in the previous section, the objective in (2.26) can be equivalently expressed as  $\|\mathbf{x} - \beta \mathbf{1}_M\|^2$ . Hence, the optimum EJT benefit vector generated by (2.26) is the Euclidean projection of the benefit vector  $\beta \mathbf{1}_M$  onto the feasible set  $\mathcal{C}$ . A subtle difference between the objectives in (2.26) and (2.25) is that, the projection in (2.26) is onto  $\mathcal{C}$ , whereas that in (2.25) is onto the intersection of  $\mathcal{C}$  with the hyperplane  $\eta(\mathbf{x}) = \sigma$ ; cf. Fig. 2.4.

To obtain further insight into the role of  $\beta$ , we note that, because of the monotonic tradeoff property, the solution of (2.21) remains unchanged if the equality constraint is replaced by the inequality  $\eta(\mathbf{x}) \geq \sigma$ . Hence,  $2\beta$  can be regarded as the Lagrange multiplier corresponding to this constraint and is, therefore, non-negative.

## 2.5 Conclusions

In this chapter, we considered multiuser resource allocations that achieve the optimal tradeoff between efficiency and fairness from the Jain's index perspective. We showed that, in general, the commonly-used  $\alpha$ -fair policy does not yield the optimal EJTs except for the two-user case. To achieve the optimal EJTs in the general case, we developed two procedures. In the first procedure, the set of feasible benefits is arbitrary, but finding the allocations that achieve the optimal EJTs involves solving

potentially difficult optimization problems. In contrast, in the second procedure, the set of feasible benefits is assumed to have a monotonic property that arises in many practical scenarios. This property is exploited to facilitate the search for allocations that achieve the optimal EJT. Our analysis is supported by illustrations, geometric interpretations, and numerical examples. Applications of the theoretical analysis presented in this chapter will be provided in the following chapter.

## Chapter 3

# Optimal Tradeoff between Sum-Rate Efficiency and Jain's Fairness Index in Resource Allocation: Applications

© 2013 IEEE. Parts of this chapter are reprinted, with permission, from:  
A. Bin Sediq, R. H. Gohary, R. Schoenen, and H. Yanıkömeroğlu, "Optimal tradeoff between sum-rate efficiency and Jain's fairness index in resource allocation," *IEEE Transactions on Wireless Communications*, vol. 12, no. 7, pp. 3496–3509, July 2013.

### 3.1 Introduction

The sufficient conditions given in Theorem 1 in Chapter 2 are quite general and can be applied to scenarios beyond those considered hereinafter. Indeed, these conditions are applicable, not only to communication systems, but also to other fields including social and economics ones.

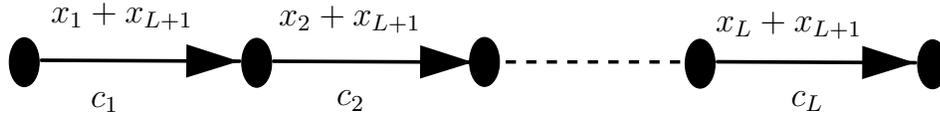
In this chapter, we demonstrate that the conditions in Theorem 1 are naturally satisfied in various resource allocation problems in communication networks. We start by demonstrating the applicability of these conditions in congestion control problems

in Section 3.2.

Another application that we study in more depth in this chapter is the scheduling of radio resources to multiple users in the downlink of a cellular network that uses orthogonal frequency division multiplexing (OFDM). Resources are divided into  $N$  (time-frequency) resource blocks (RBs) [23], and the goal is to allocate these RBs to  $M$  users in a way that is both “efficient and fair”. To facilitate the analysis, a full-buffer traffic model is used, i.e., the users are assumed to have an infinite backlog, whereupon data is always available for transmissions. In this Chapter, we assume that each BS makes its scheduling decisions independent of other BSs in the network, i.e., we do not consider ICIC. A thorough treatment of using ICIC for efficient and fair multi-cell radio resource allocation will be provided in Chapter 4. We consider the cases of both quasi-static and ergodic time-varying channels.

We analyze the case of scheduling in quasi-static channels in Section 3.3. For quasi-static channels, we consider scheduling with and without time-sharing. In the case of time-sharing, the scheduling variables are continuous and the corresponding set of feasible benefit vectors,  $\mathcal{C}$ , satisfies the conditions of Theorem 1. In contrast, in the case without time-sharing, the scheduling variables are discrete and  $\mathcal{C}$  does not satisfy the conditions of Theorem 1. The analysis of the case of scheduling in quasi-static channels is supported by numerical results in Section 3.4.

In Section 3.5, we analyze the case of scheduling in ergodic time-varying channels. In this case, time-sharing is not plausible and the scheduling variables are discrete. In spite of that, the corresponding set of feasible benefit vectors,  $\mathcal{C}$ , can be shown to satisfy the conditions of Theorem 1. We support our analysis of the case of scheduling in time-varying channels by simulation results in Section 3.6.



**Figure 3.1:** A linear network that consists of  $L$  links and  $L + 1$  users. User  $m$ ,  $1 \leq m \leq L$ , uses only the  $m^{\text{th}}$  link, while user  $L + 1$  uses all the  $L$  links.

## 3.2 Congestion Control

In congestion control in elastic traffic communication networks [3, 13],  $M$  users share  $L$  finite-capacity links and the goal is to obtain an efficient and fair benefit vector  $\mathbf{x}$ , which represents the users' rates. The set of feasible rates in this case is given by  $\mathcal{C} = \{\mathbf{x} | \mathbf{A}\mathbf{x} \preceq \mathbf{c}, \mathbf{0}_M \preceq \mathbf{x}\}$ , where the  $\ell$ -th entry of  $\mathbf{c} \in \mathbb{R}_+^L$  is the capacity of link  $\ell$ , for  $\ell = 1, \dots, L$ , and  $\mathbf{A}$  is a matrix with binary entries that represent the assignments of users to links. In this case, the set  $\mathcal{C}$  is a convex polyhedron [17, p. 31] containing  $\mathbf{0}_M$ , and thereby satisfying the conditions of Theorem 1 with  $x_{\min} = 0$ . Hence,  $\mathcal{C}$  satisfies the monotonic tradeoff property and Procedure 2 can be efficiently used to find all the optimal EJT rate vectors.

### 3.2.1 Special Case: Linear Networks

Consider a network with  $L$  links, each with a capacity of  $c_\ell$ . These  $L$  links are shared by  $M = L + 1$  users. User  $m$ ,  $1 \leq m \leq L$ , uses only the  $m^{\text{th}}$  link, while user  $L + 1$  uses all the  $L$  links. This linear network is depicted in Fig. 3.1. The feasible set of rates is given by

$$\mathcal{C} = \{\mathbf{x} | x_m + x_{L+1} \leq c_m, 1 \leq m \leq L; x_m \geq 0, 1 \leq m \leq L + 1\}. \quad (3.1)$$

A linear network with uniform capacity is one which has equal capacities of one unit, i.e.,  $c_\ell = 1, \forall \ell \in [1, L]$  [24].

In the following, we investigate the relationship between the  $\alpha$ -fair policy and the optimal EJT for a linear network with uniform capacity. The main result in this case is stated in the following proposition:

**Proposition 2.** *Given a linear network with uniform capacity and  $L$  links, the  $\alpha$ -fair rate vector  $\mathbf{x}_\alpha^*$  generated by (2.2) achieves the optimal EJT, for any  $\alpha \in [0, \infty)$ .*

*Proof.* As shown in [24, 25], the  $\alpha$ -fair rate vector for the considered linear network with uniform capacity assumes the following form:

$$x_{\alpha,m}^* = \begin{cases} \frac{L^{\frac{1}{\alpha}}}{1 + L^{\frac{1}{\alpha}}}, & 1 \leq m \leq L, \\ \frac{1}{1 + L^{\frac{1}{\alpha}}}, & m = L + 1. \end{cases} \quad (3.2)$$

In the following, we show that the  $\alpha$ -fair rate vector achieves the optimal EJT. To do so, we start by noting that the set  $\mathcal{C}$  is a convex polyhedron [17, p. 31] containing  $\mathbf{0}_M$ , and thereby satisfying the conditions of Theorem 1 with  $x_{\min} = 0$ . Hence,  $\mathcal{C}$  satisfies the monotonic tradeoff property and Procedure 2 can be used to find all the optimal EJT rate vectors. According to Procedure 2, to prove that the rate vector  $\mathbf{x}_\alpha^*$  given by (3.2) is an optimal EJT rate vector, it suffices to show that  $\mathbf{x}_\alpha^*$  is also the optimum solution of the following optimization problem:

$$\min_{\eta(\mathbf{x}) = \frac{1+L^{1+\frac{1}{\alpha}}}{1+L^{\frac{1}{\alpha}}}, \mathbf{x} \in \mathcal{C}} \|\mathbf{x}\|^2 \quad (3.3)$$

In order to solve the optimization problem given by (3.3), we begin by noting that it is a convex optimization problem since it involves minimizing a convex function over a convex set. Moreover, since the constraints are linear inequalities, Slater's condition is satisfied as long as the problem is feasible, which is always the case [17, p. 227]. As a result, the Karush-Kuhn-Tucker (KKT) conditions are necessary and sufficient for

optimality. To express the KKT conditions, we define the Lagrangian corresponding to the optimization problem (3.3) as

$$L(\mathbf{x}, \lambda, \mathbf{v}, u) = \|\mathbf{x}\|^2 + \sum_{m=1}^L \lambda_m (x_m + x_{L+1} - 1) - \sum_{m=1}^{L+1} v_m x_m + u \left( \eta(\mathbf{x}) - \frac{1 + L^{1+\frac{1}{\alpha}}}{1 + L^{\frac{1}{\alpha}}} \right), \quad (3.4)$$

where  $\lambda_m$  is the Lagrange multiplier associated with the  $m^{\text{th}}$  inequality constraint,  $x_m + x_{L+1} \leq 1$ ,  $v_m$  is Lagrange multiplier associated with  $m^{\text{th}}$  nonnegativity inequality constraint,  $x_m \geq 0$ , and  $u$  is the Lagrange multipliers associated with the efficiency equality constraint,  $\eta(\mathbf{x}) = \frac{1+L^{1+\frac{1}{\alpha}}}{1+L^{\frac{1}{\alpha}}}$ . The optimal solution of (3.3),  $\mathbf{x}^*$ , and the optimal Lagrange multipliers  $\lambda^*$ ,  $\mathbf{v}^*$ , and  $u^*$  are characterized by the KKT conditions given below

$$\text{Stationarity: } \frac{\partial}{\partial x_m} L(\mathbf{x}, \lambda, \mathbf{v}, u) = \begin{cases} 2x_m^* + \lambda_m^* - v_m^* + u^* = 0, & 1 \leq m \leq L, \\ 2x_{L+1}^* + \sum_{m=1}^L \lambda_m^* - v_{L+1}^* + u^* = 0, & m = L + 1. \end{cases}$$

$$\text{Primal feasibility: } \mathbf{x}^* \in \mathcal{C}, \eta(\mathbf{x}) = \frac{1+L^{1+\frac{1}{\alpha}}}{1+L^{\frac{1}{\alpha}}}.$$

$$\text{Dual feasibility: } \lambda_m^* \geq 0, \quad 1 \leq m \leq L,$$

$$v_m^* \geq 0, \quad 1 \leq m \leq L + 1.$$

$$\text{Complimentray slackness: } \lambda_m^* (x_m^* + x_{L+1}^* - 1) = 0, \quad 1 \leq m \leq L,$$

$$v_m^* x_m^* = 0, \quad 1 \leq m \leq L + 1.$$

(3.5)

By solving the equations given in (3.5), it can be shown that optimal primal and dual

solutions assume the following form:

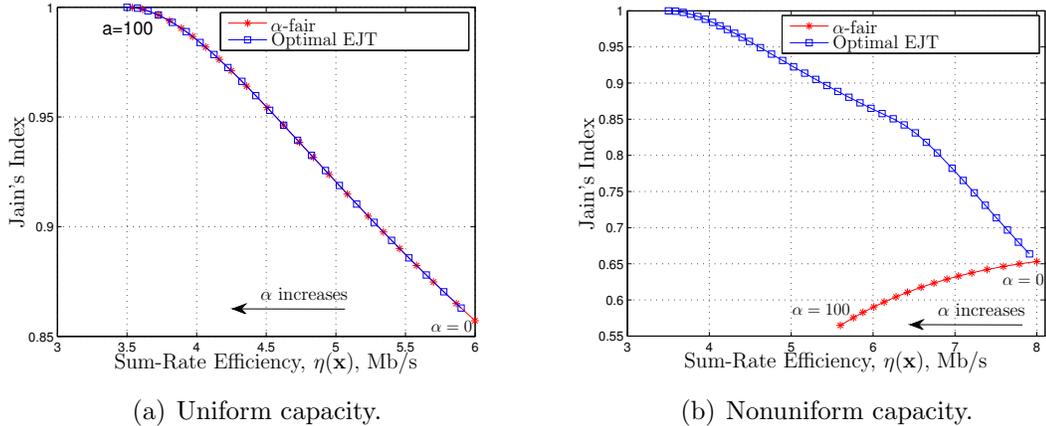
$$\begin{aligned}
\text{Primal optimal solution: } x_m^* = x_{\alpha,m}^* &= \begin{cases} \frac{L^{\frac{1}{\alpha}}}{1 + L^{\frac{1}{\alpha}}}, & 1 \leq m \leq L, \\ \frac{1}{1 + L^{\frac{1}{\alpha}}}, & m = L + 1. \end{cases} \\
\text{Dual optimal solution: } \lambda_m^* &= 2 \frac{-1 + L^{\frac{1}{\alpha}}}{(1 + L^{\frac{1}{\alpha}})(L - 1)}, \\
v_m^* &= 0, \\
u^* &= -2 \frac{-1 + L^{1+\frac{1}{\alpha}}}{(1 + L^{\frac{1}{\alpha}})(L - 1)}.
\end{aligned} \tag{3.6}$$

Thus,  $\mathbf{x}_\alpha^*$  achieves the optimal EJT. □

Unlike Proposition 1 provided in Chapter 2, where we proved that the  $\alpha$ -fair tradeoff policy yields the optimal EJT for an arbitrary set  $\mathcal{C}$  and  $M = 2$  users, Proposition 2 shows that for a linear network with  $L$  links,  $M = L + 1$  users, and uniform capacity, the  $\alpha$ -fair policy yields the optimal EJT. However, this result does not necessarily carry over to linear networks with nonuniform capacity. To show this, we consider linear networks with uniform and nonuniform capacity. In both cases, the considered networks consist of  $L = 6$  links and  $M = L + 1 = 7$  users. For the uniform case, the capacity of the links are given by  $c_\ell = 1$  Mb/s  $\forall \ell \in [1, L]$ . To introduce nonuniformity among the capacities of the links, we only change the capacity of the first link to 3 Mb/s, i.e., the capacity of the  $\ell$ -th link for the nonuniform case is given by

$$c_\ell = \begin{cases} 3 \text{ Mb/s}, & \ell = 1, \\ 1 \text{ Mb/s}, & 2 \leq \ell \leq L. \end{cases}$$

The comparisons between the optimal EJT and the EJT achieved by the  $\alpha$ -fair policy for the considered linear networks with uniform and nonuniform capacity are



**Figure 3.2:** A comparison between the EJT's achieved by the optimal and  $\alpha$ -fair policies in a linear network.

shown in Fig. 3.2(a) and Fig. 3.2(b), respectively. The optimal EJT was obtained using Procedure 2 and the EJT achieved by the  $\alpha$ -fair policy was obtained by solving (2.2). The convex optimization problems underlying Procedure 2 and the  $\alpha$ -fair policy were solved using ‘fmincon’, which is available in the MATLAB optimization toolbox. In agreement with Proposition 2, Fig. 3.2(a) shows that the  $\alpha$ -fair tradeoff policy achieves the optimal EJT for the case of uniform capacity. However, for the case of nonuniform capacity depicted in Fig. 3.2(b), it can be seen that the tradeoff generated by the  $\alpha$ -fair tradeoff policy is far from the one generated by the optimal EJT policy. Indeed, this example reiterates the fact that increasing  $\alpha$  improves the fairness from  $\alpha$ -fairness perspective but can result in strict decrease in Jain’s fairness index, as shown in Fig. 3.2(b).<sup>1</sup> This example also demonstrates that, unlike the optimal EJT policy, the  $\alpha$ -fair tradeoff policy guarantees Pareto optimality in the user rates but it does not guarantee Pareto optimality in efficiency and Jain’s fairness index.

<sup>1</sup>In this example, the set of feasible rates does not satisfy the solidarity property (cf. [22, Section IV]), hence the max-min fair allocation, achieved by setting  $\alpha \rightarrow \infty$ , is not guaranteed to result in equal rates for all users.

### 3.3 Radio Resource Allocations in Quasi-Static Channels: Analysis

To elaborate on the motivation for considering quasi-static channels, we note that these channels are frequently encountered in various practical communication scenarios. We mention two examples for brevity:

- a. Digital Subscriber Line (DSL) wireline communications: In these systems, the quality of channels depends on electromagnetic coupling between lines and on the distance from the central office to subscribing units. Such channels remain constant for numerous time slots and can be assumed quasi-static for essentially all practical considerations.
- b. Relay assisted cellular systems: In these systems, the base station (BS) communicates with the user terminal (UT) in two hops: one hop from the BS to a fixed relay and one hop from the relay to the UT. When the relay is close to the UT, the BS-UT communication is constrained by the channel from the BS to the relay. Because both the BS and the relay are fixed, the channel between them is dominated by pathloss and shadowing. Hence, this channel remains constant for numerous time slots and can be assumed quasi-static for many practical considerations.

Under quasi-static channel conditions and given modulation and coding schemes, the data rate of each user  $m \in \mathcal{M} \triangleq \{1, \dots, M\}$  on RB  $n \in \mathcal{N} \triangleq \{1, \dots, N\}$ , which we denote by  $r_{m,n}$ , is a deterministic quantity known to the transmitter. The objective of the transmitter is to determine a fair RB allocation that ensures efficient communication of data to the users. To achieve this goal, let  $\rho_{m,n} \in [0, 1]$  be a scheduling variable that assigns RB  $n$  to user  $m$  for a fraction  $\rho_{m,n}$  of the signalling interval [26]. At each time instant, each RB is used by at most one user, and thus  $\sum_{m=1}^M \rho_{m,n} \leq 1$ . The total data rate (benefit) of user  $m$  is given by  $x_m = \sum_{n=1}^N \rho_{m,n} r_{m,n}$  and the efficiency

of the network is given by the total sum-rate, which is given by  $\eta(\mathbf{x}) = \sum_{m=1}^M x_m$ .

The set of achievable rates (benefits) of the users is given by

$$\mathcal{C} = \left\{ \mathbf{x} \mid x_m = \sum_{n=1}^N \rho_{m,n} r_{m,n}, \sum_{m=1}^M \rho_{m,n} \leq 1, \rho_{m,n} \in [0, 1], x_m \geq x_{\min}, m \in \mathcal{M}, n \in \mathcal{N} \right\}, \quad (3.7)$$

where  $x_{\min} \geq 0$  represents a feasible threshold on the minimum rate that must be delivered to each user. Using this description, the goal of the transmitter can be cast as to determine the set  $\{\rho_{m,n}\}$  that results in rate vectors  $\mathbf{x}$  that span the optimal EJT. This goal can be achieved by invoking the results of Theorem 1. In particular, we note that the set  $\mathcal{C}$  in (3.7) is convex and contains the vector  $x_{\min} \mathbf{1}_M$ . Hence, the conditions of Theorem 1 are satisfied and  $\mathcal{C}$  possesses the monotonic tradeoff property. Based on this observation, Procedure 2 will be used in Section 3.4 to obtain  $\{\rho_{m,n}\}$  that achieve every point on the optimal EJT.

When the RBs are not time-shared,  $\{\rho_{m,n}\}$  assume binary values (i.e.,  $\rho_{m,n} \in \{0, 1\}$ ), resulting in the set  $\mathcal{C}$  being non-convex. In this case,  $\mathcal{C}$  may not possess the monotonic tradeoff property and Procedure 1 can be used to obtain the optimal  $\{\rho_{m,n}\}$ . Procedure 1 is significantly more computationally-demanding in comparison with Procedure 2, which is used when time-sharing is allowed and  $\{\rho_{mn}\}$  are continuous.

For the sake of presentation, the optimal EJT scheduler is defined as the scheduler that finds the scheduling variables,  $\{\rho_{m,n}\}$ , based on the optimal EJT policy. Similarly, the  $\alpha$ -fair scheduler is defined as the scheduler that finds the scheduling variables,  $\{\rho_{m,n}\}$ , based on the  $\alpha$ -fair tradeoff policy.

### 3.3.1 Special Case: Scheduling in Frequency-Flat Fading with Time-Sharing

In the following, we analyze the performance of the  $\alpha$ -fair scheduler and the optimal EJT scheduler in quasi-static frequency-flat fading channels with time-sharing. In particular, for both schedulers, closed-form expressions of the data rate delivered to each user are derived and useful insights are drawn from these expressions. In frequency-flat fading channels, the data rate of each user is the same on all RBs, i.e.,  $r_{m,n} = r_{m\tilde{n}} = r_m, \forall n, \tilde{n} \in \mathcal{N}, \forall m \in \mathcal{M}$ .

#### $\alpha$ -fair Based Scheduling

In quasi-static frequency-flat fading channel, the  $\alpha$ -fair rate vector,  $\mathbf{x}_\alpha^*$ , assumes the following form [27]

$$x_{\alpha,m}^* = \frac{r_m^{\frac{1}{\alpha}}}{\sum_{m=1}^M r_m^{\frac{1-\alpha}{\alpha}}} N \quad (3.8)$$

In the following, we examine an important special case of  $\alpha$ -fair scheduling, which is proportional-fair scheduling that corresponds to  $\alpha = 1$ . In this case, the proportional-fair rates are given by

$$x_{\alpha=1,m}^* = \frac{N}{M} r_m. \quad (3.9)$$

According to (3.9), a proportional-fair scheduler is a resource-fair scheduler, i.e., it allocates equal resources to all users regardless of the rates they can achieve in each RB,  $\{r_m\}$  [27, 28]. This implies that the resulted efficiency and resulted Jain's index are given by  $\eta(\mathbf{x}_{\alpha=1}^*) = \frac{N}{M} \sum_{m=1}^M r_m$  and  $J(\mathbf{x}_{\alpha=1}^*) = \frac{(\sum_{m=1}^M r_m)^2}{M \sum_{m=1}^M r_m^2}$ , respectively. As a result, the EJT achieved by the proportional-fair scheduler is highly dependant on the users' rates  $\{r_m\}$ . To elaborate, we consider two extreme cases, where in the first case, the proportional-fair scheduler achieves the best efficiency and fairness, while in

the second case, it achieves asymptotically the least fair and only  $\frac{1}{M}$  of the maximum achievable efficiency.

Consider the case where all users have the same rates, i.e.,  $r_m = r, \forall m \in \mathcal{M}$ . In this case, the most efficient and most fair solution is to allocate resources equally to all users, which is achieved by the proportional-fair scheduler. On the contrary, the proportional-fair scheduler does not perform as good when users' rates vary significantly. To illustrate this, we consider the case where the users' rates are given by a geometric sequence as  $\{r_m = \epsilon^{m-1}, m \in \mathcal{M}\}$ , where  $\epsilon$  is an arbitrarily small positive number, i.e.,  $0 < \epsilon \ll 1$ . In this case, the resulted efficiency and Jain's index are given by

$$\eta(\mathbf{x}_{\alpha=1}^*) = \frac{N}{M} \sum_{m=1}^M r_m = \frac{N}{M} \sum_{m=1}^M \epsilon^{m-1} = \frac{1 - \epsilon^M}{1 - \epsilon} \stackrel{\epsilon \rightarrow 0}{\approx} \frac{N}{M}, \quad (3.10)$$

$$J(\mathbf{x}_{\alpha=1}^*) = \frac{\left(\sum_{m=1}^M r_m\right)^2}{M \sum_{m=1}^M r_m^2} = \frac{\left(\sum_{m=1}^M \epsilon^{m-1}\right)^2}{M \sum_{m=1}^M \epsilon^{2(m-1)}} = \frac{\frac{(1-\epsilon^M)^2}{(1-\epsilon)^2}}{M \frac{1-\epsilon^{2M}}{1-\epsilon^2}} = \frac{(1-\epsilon^M)(1+\epsilon)}{M(1+\epsilon^M)(1-\epsilon)} \stackrel{\epsilon \rightarrow 0}{\approx} \frac{1}{M}. \quad (3.11)$$

Examining the asymptotic expressions given by (3.10) and (3.11) reveals that proportional-fair scheduling leads to the least-fair solution as measured by the Jain's index (recall that  $J(\mathbf{x}_{\alpha=1}^*) \in [\frac{1}{M}, 1]$ ). Moreover, the efficiency achieved by such allocation is only  $\frac{1}{M}$  of the maximum feasible efficiency achieved by assigning all resources to the best user, which is user 1 in this case. Hence, in this particular user configuration, the proportional-fair scheduler is not a good choice from EJT perspective as it yields both low efficiency and low fairness.

## Optimal EJT Scheduling

As discussed earlier, Procedure 2 can be used to find allocations that yield the optimal EJT. In this procedure, the optimization problem in (2.21) is solved for

various choices of  $\sigma$ . For each value of  $\sigma$ , the problem in (2.21) involves a constraint on the sum-rate efficiency. Unfortunately, such a constraint complicates the derivation of a closed-form solution. To circumvent this difficulty, we use the equivalent alternate formulation of (2.21) given in (2.26). Using this formulation, the optimal EJT rate vector, for a given  $\beta$ , is given by

$$\mathbf{x}_\beta^* = \min_{\mathbf{x} \in \mathcal{C}} \|\mathbf{x}\|^2 - 2\beta\eta(\mathbf{x}), \quad (3.12)$$

where

$$\mathcal{C} = \left\{ \mathbf{x} \mid \mathbf{x}^T \mathbf{r}^{-1} \leq N, \mathbf{x} \succeq x_{\min} \mathbf{1}_M, m \in \mathcal{M} \right\}, \quad (3.13)$$

and  $\mathbf{r}^{-1}$  is a length- $M$  vector given by  $\mathbf{r}^{-1} = [r_1^{-1}, \dots, r_M^{-1}]^T$ .

In order to solve the optimization problem given by (3.12), we begin by noting that (3.12) is a convex optimization problem since both the objective function and the feasible set are convex. Moreover, since the constraints that define  $\mathcal{C}$  are linear inequalities, Slater's condition is satisfied for non-empty  $\mathcal{C}$  [17, p. 227]. As a result, the KKT conditions are necessary and sufficient for optimality. To express the KKT conditions, we define the Lagrangian corresponding to the optimization problem (3.12) as

$$L(\mathbf{x}, \lambda, \mathbf{v}) = \|\mathbf{x}\|^2 - 2\beta\eta(\mathbf{x}) + \lambda(\mathbf{x}^T \mathbf{r}^{-1} - N) - \mathbf{v}^T(\mathbf{x} - x_{\min} \mathbf{1}_M), \quad (3.14)$$

where  $\lambda$  and  $\mathbf{v}$  are the Lagrange multipliers associated with the constraints  $\mathbf{x}^T \mathbf{r}^{-1} \leq N$  and  $\mathbf{x} \succeq \mathbf{0}$ , respectively. For a given  $\beta$ , the optimal solution to (3.12),  $\mathbf{x}_\beta^*$ , and the optimal Lagrange multipliers,  $\lambda_\beta^*$  and  $\mathbf{v}_\beta^*$ , are characterized by the KKT conditions

given below

$$\text{Stationarity: } \nabla_{\mathbf{x}} L(\mathbf{x}_\beta^*, \lambda_\beta^*, \mathbf{v}_\beta^*) = 2\mathbf{x}_\beta^* - 2\beta\mathbf{1}_M + \lambda_\beta^* \mathbf{r}^{-1} - \mathbf{v}_\beta^* = \mathbf{0}.$$

$$\text{Primal feasibility: } \mathbf{x}_\beta^{*T} \mathbf{r}^{-1} \leq N, \mathbf{x} \succeq x_{\min} \mathbf{1}_M. \quad (3.15)$$

$$\text{Dual feasibility: } \lambda_\beta^* \geq 0, \mathbf{v}_\beta^* \succeq \mathbf{0}_M.$$

$$\text{Complimentary slackness: } \lambda_\beta^* (\mathbf{x}_\beta^{*T} \mathbf{r}^{-1} - N) = 0, v_{\beta,m}^* (x_{\beta,m}^* - x_{\min}) = 0.$$

Using the stationarity and complimentary slackness conditions, it is not difficult to show that the solution to the KKT conditions in (3.15) is given by

$$x_{\beta,m}^* = \max(x_{\min}, \beta - 0.5\lambda_\beta^* r_m^{-1}), \quad (3.16)$$

$$v_{\beta,m}^* = 2x_{\beta,m}^* - 2\beta + \lambda_\beta^* r_m^{-1}. \quad (3.17)$$

To compute  $\lambda_\beta^*$ , we note that  $\mathbf{x}_\beta^{*T} \mathbf{r}^{-1} \leq N$  is satisfied with equality since  $\mathbf{x}_\beta^*$  is an optimal EJT rate vector. As a result,  $\lambda_\beta^*$  can be computed by solving the following equation

$$\sum_{m=1}^M r_m^{-1} \max(x_{\min}, \beta - 0.5\lambda_\beta^* r_m^{-1}) = N. \quad (3.18)$$

To facilitate solving (3.18), we define  $\hat{\mathcal{M}}_\beta$  as the set of users who receives exactly the minimum rate requirement, i.e.,  $\hat{\mathcal{M}}_\beta = \{m | x_{\beta,m}^* = x_{\min}\}$ . Using the definition of  $\hat{\mathcal{M}}_\beta$ , it can be readily verified that the solution to (3.18) assumes the following form

$$\lambda_\beta^* = 2 \left( (\eta(\mathbf{r}^{-1}) - \eta(\hat{\mathbf{r}}_\beta^{-1})) x_{\min} + \eta(\hat{\mathbf{r}}_\beta^{-1}) \beta - N \right) M \frac{J(\hat{\mathbf{r}}_\beta^{-1})}{\eta^2(\hat{\mathbf{r}}_\beta^{-1})} \quad (3.19)$$

where  $\hat{\mathbf{r}}_\beta^{-1}$  is a vector of non-negative real entries  $\{\hat{r}_{\beta,m}^{-1}\}_{m=1}^M$  given by

$$\hat{r}_{\beta,m}^{-1} = \begin{cases} r_m^{-1}, & m \notin \hat{\mathcal{M}}_\beta, \\ 0, & m \in \hat{\mathcal{M}}_\beta. \end{cases} \quad (3.20)$$

Thus, obtaining  $\lambda_\beta^*$  requires calculating  $\hat{\mathcal{M}}_\beta$ . In order to compute  $\hat{\mathcal{M}}_\beta$ , we note that the left-hand side of (3.18) is monotonically decreasing in the cardinality of  $\hat{\mathcal{M}}_\beta$ . Moreover, the largest decrease in the left-hand side of (3.18) occurs by including the user with highest  $r_m^{-1}$  (i.e., smallest  $r_m$ ) in  $\hat{\mathcal{M}}_\beta$ . Based on these observations, we devise Algorithm 3 to compute  $\hat{\mathcal{M}}_\beta$ . It is not difficult to show that Algorithm 3 given below has a worst-case complexity of  $O(M^2)$ .

---

**Algorithm 3** Finding  $\hat{\mathcal{M}}_\beta = \{m | x_{\beta,m}^* = x_{\min}\}$

---

**Input:**  $\mathbf{r}^{-1}$  and  $M$

**Output:**  $\hat{\mathcal{M}}_\beta$

- 1: Initialize  $\hat{\mathcal{M}}_\beta = \Phi$
  - 2: **for**  $p = 1 : M - 1$  **do**
  - 3:   Calculate  $\hat{\mathbf{r}}_\beta^{-1}$  using (3.20) and  $\lambda_\beta^*$  using (3.19).
  - 4:   **if**  $\sum_{m=1}^M r_m^{-1} \max(x_{\min}, \beta - 0.5\lambda_\beta^* r_m^{-1}) = N$  and  $\lambda_\beta^* > 0$  **then**
  - 5:     return  $\hat{\mathcal{M}}_\beta$ .
  - 6:   **else**
  - 7:      $\hat{\mathcal{M}}_\beta = \hat{\mathcal{M}}_\beta \cup \hat{m}$ , where  $\hat{m} = \arg \min_{m \notin \hat{\mathcal{M}}_\beta} r_m$
  - 8:   **end if**
  - 9: **end for**
  - 10: return  $\hat{\mathcal{M}}_\beta$ .
- 

Once  $\hat{\mathcal{M}}_\beta$  is computed using Algorithm 3,  $\lambda_\beta^*$  can be calculated using (3.19). Substituting (3.19) in (3.16), the optimal solution  $x_{\beta,m}^*$  can be expressed as

$$x_{\beta,m}^* = \max \left( x_{\min}, \beta - \left( (\eta(\mathbf{r}^{-1}) - \eta(\hat{\mathbf{r}}_\beta^{-1})) x_{\min} + \eta(\hat{\mathbf{r}}_\beta^{-1}) \beta - M \right) M \frac{J(\hat{\mathbf{r}}_\beta^{-1})}{\eta^2(\hat{\mathbf{r}}_\beta^{-1})} r_m^{-1} \right), \quad (3.21)$$

An important insight that can be drawn from the structure of the optimum solution given by (3.21) is how the optimal EJT scheduler handles potentially disadvantageous users. In particular, in the following we show that increasing  $\beta$  increases the resulting efficiency by increasing the number of users in  $\hat{\mathcal{M}}_\beta$  who are served with just the minimum rate requirement  $x_{\min}$  and allocating the remaining resources fairly among the more advantageous users. To show that, let  $m_{\min}$  denote the index of the user with the lowest  $r_m$  and it is not included in  $\hat{\mathcal{M}}_\beta$ , i.e.,  $m_{\min} \in \arg \min_{m \notin \hat{\mathcal{M}}_\beta} r_m$ . We show that  $x_{\beta+\Delta, m_{\min}}^* \leq x_{\beta, m_{\min}}^*$  for sufficiently small  $\Delta > 0$ . For sufficiently small  $\Delta > 0$ ,  $\hat{\mathcal{M}}_{\beta+\Delta} = \hat{\mathcal{M}}_\beta$ . Hence, using (3.21), we can write

$$x_{\beta+\Delta, m_{\min}}^* - x_{\beta, m_{\min}}^* = \Delta \left( 1 - M \frac{J(\hat{\mathbf{r}}_\beta^{-1})}{\eta(\hat{\mathbf{r}}_\beta^{-1})} r_{m_{\min}}^{-1} \right) \leq \Delta \left( 1 - \frac{M r_{m_{\min}}^{-1}}{\eta(\hat{\mathbf{r}}_\beta^{-1})} \right) \leq 0, \quad (3.22)$$

where the first inequality follows from the fact that  $J(\hat{\mathbf{r}}_\beta^{-1}) \leq 1$  and the second inequality follows from the fact that  $r_{m_{\min}}^{-1} \geq r_m^{-1}, \forall m \notin \hat{\mathcal{M}}_\beta$ . Note that the inequality in (3.22) holds with equality if and only if  $r_{m_{\min}}^{-1} = r_m^{-1}, \forall m \notin \hat{\mathcal{M}}_\beta$ . Hence, we proved that as  $\beta$  increases,  $x_{\beta, m_{\min}}^*$  decreases until  $x_{\beta, m_{\min}}^* = x_{\min}$ . Thus, increasing  $\beta$  can potentially increase the cardinality of  $\hat{\mathcal{M}}_\beta$ .

On the other hand, a similar proof can be constructed to show that as the parameter  $\beta$  decreases, less and less users are included in  $\hat{\mathcal{M}}_\beta$ . This leads to fairer solutions as more and more users are served with rates that are strictly higher than  $x_{\min}$ .

### 3.4 Radio Resource Allocations in Quasi-Static Channels: Numerical Results

In this section, we compare the EJT achieved by the optimal EJT scheduler and the  $\alpha$ -fair one in quasi-static channels through numerical examples. We start by considering a special case of frequency-flat fading and then we consider the general

case of frequency-selective fading.

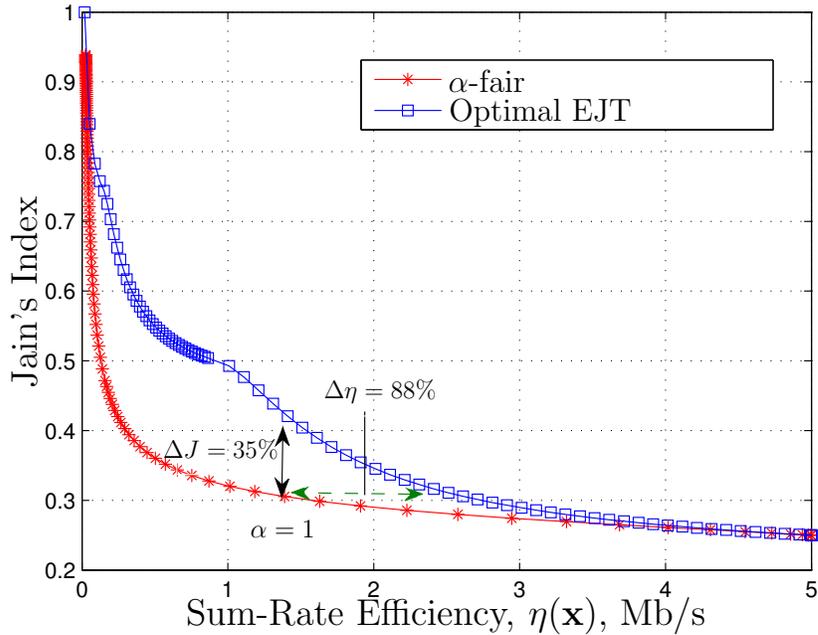
### 3.4.1 Frequency-Flat Fading

We consider one realization of a quasi-static frequency-flat channel in a network with  $M = 4$  users and  $N = 5$  RBs. As an example, we assume that the rate matrix is given by a geometric sequence as  $\mathbf{r} = [r_{m,n} = 1000 \times 0.1^{m-1} \text{ Kb/s}]$ , i.e.,

$$\mathbf{r} = \begin{bmatrix} 1000 & 1000 & 1000 & 1000 & 1000 \\ 100 & 100 & 100 & 100 & 100 \\ 10 & 10 & 10 & 10 & 10 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix}. \quad (3.23)$$

For simplicity, we assume there is no minimum rate guarantee, i.e.,  $x_{\min} = 0$ , and we consider only the case of time-sharing. The comparison between the optimal EJT and the EJT achieved by the  $\alpha$ -fair scheduler for the considered example is shown in Fig. 3.3. The optimal EJT was obtained using Procedure 2 and the EJT trade-off achieved by the  $\alpha$ -fair scheduler was obtained by solving (2.2). The solutions of the convex optimization problems underlying Procedure 2 are obtained using the closed-form expression given in (3.21) while the solutions of the convex optimization problems underlying the  $\alpha$ -fair scheduler were obtained using the closed-form expression given in (3.8) which was derived in [27]. From this figure, it can be seen that the EJT generated by the optimal EJT scheduler is significantly better than that generated by the  $\alpha$ -fair one. In particular, compared to the proportional-fair case ( $\alpha = 1$ ), the optimal EJT scheduler can provide a gain of 88% in efficiency for the same Jain's index or a gain of 35% in Jain's index for the same efficiency. Such large gains agree with the insights drawn from the analysis of proportional-fairness in Section 3.3.1.

To gain further insights on the difference between the optimal EJT scheduler and the  $\alpha$ -fair one, we plot in Fig. 3.4 the users' rates at an efficiency of 1 Mb/s for both schedulers. It can be seen that for the case of the  $\alpha$ -fair scheduler, all users have non-zero rates, although there are large variations between these rates. Such large variations result in a low Jain's fairness index of 0.32. On the other hand, for the case of the optimal EJT scheduler, only two users have non-zero rates with very close values. This results in the maximum Jain's index that can be achieved at an efficiency of 1 Mb/s for the considered example, which is 0.49. This comparison shows a fundamental difference between the  $\alpha$ -fair scheduler and the optimal EJT one on the way they achieve high efficiency while serving a mixture of advantageous user, i.e., users with large  $r_{m,n}$ , and disadvantageous users, i.e., users with small  $r_{m,n}$ . At high efficiencies, the  $\alpha$ -fair scheduler tends to provide disadvantageous users with very small non-zero rates and advantageous users with large rates. On the other hand, the optimal EJT scheduler tends to provide the disadvantageous users with just the minimum required rate ( $x_{\min}$ ) and allocate the remaining resources fairly to the advantageous users. This observation agrees with the insights drawn from the analysis of the optimal EJT in frequency-flat fading in Section 3.3.1. Indeed, we will demonstrate in 3.6 that this observation also holds for the case of resource allocation in ergodic time-varying frequency-selective channels.



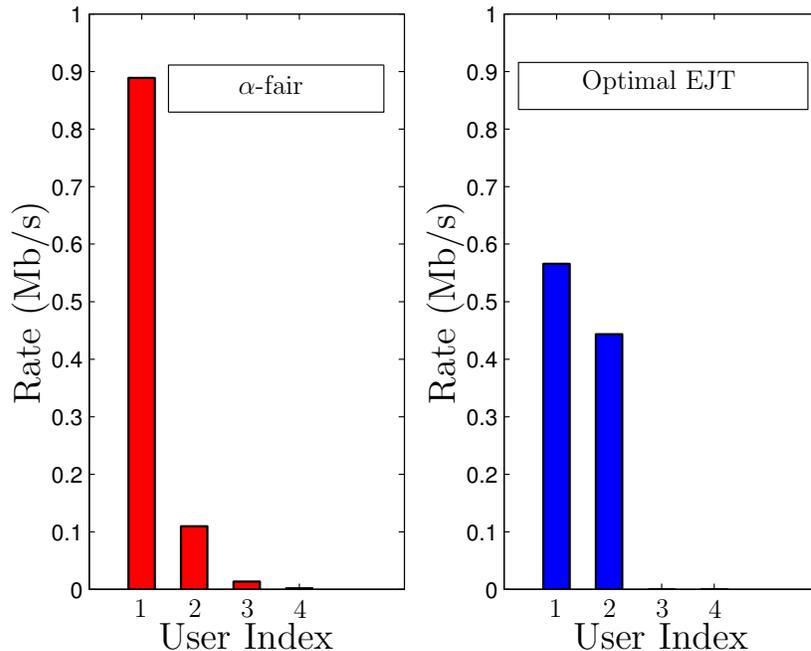
**Figure 3.3:** A comparison between the EJT's achieved by the optimal and  $\alpha$ -fair schedulers in frequency-flat channels with time-sharing.

### 3.4.2 Frequency-Selective Fading

We consider one realization of a quasi-static network with  $M = 4$  users and  $N = 5$  RBs. As an example, we assume that the rate matrix  $\mathbf{r} = [r_{m,n}]$  is given by

$$\mathbf{r} = \begin{bmatrix} 544 & 648 & 807 & 544 & 722 \\ 388 & 92 & 223 & 388 & 56 \\ 35 & 544 & 35 & 722 & 56 \\ 35 & 56 & 35 & 92 & 35 \end{bmatrix}. \quad (3.24)$$

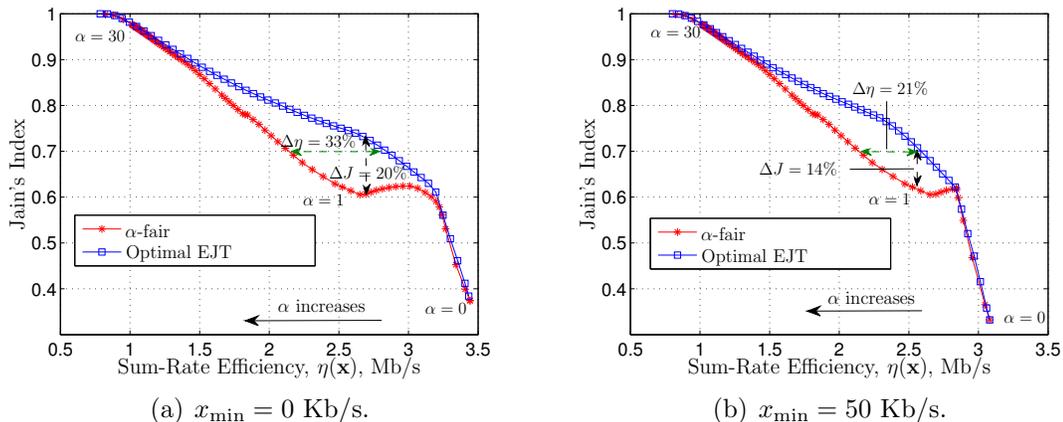
The rates in this matrix are given in Kb/s and were obtained from simulating a practical scenario based on the Long Term Evolution (LTE) standard [29]. In the considered scenario, users 1 and 2 are closer to the BS than users 3 and 4, and the



**Figure 3.4:** A comparison between the users’ rates achieved by the optimal EJT and  $\alpha$ -fair schedulers in frequency-flat channels with time-sharing at an efficiency of 1 Mb/s.

wireless channels are quasi-static frequency-flat and Rayleigh fading on each RB.

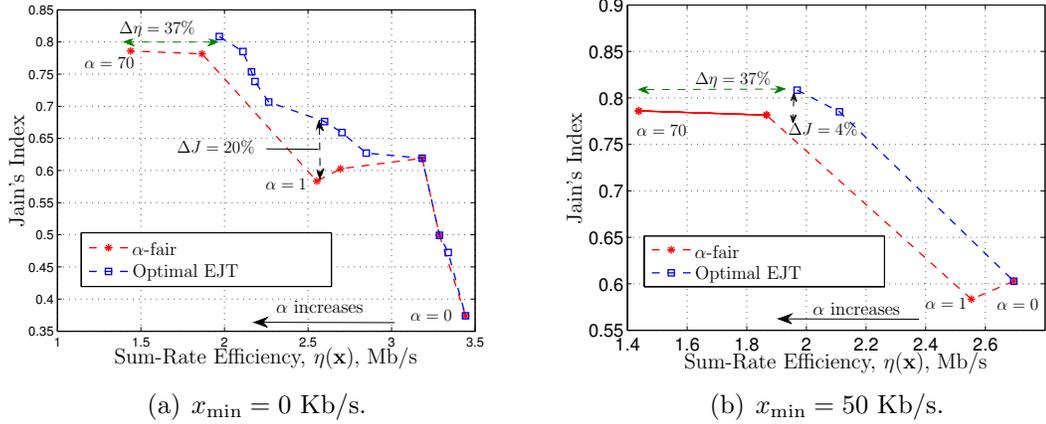
The comparisons between the optimal EJT and the tradeoff achieved by the  $\alpha$ -fair scheduler for the case when the RBs can be time-shared among users are shown in Fig. 3.5(a) and Fig. 3.5(b). In Fig. 3.5(a), it is assumed that there is no minimum rate guarantee, i.e.,  $x_{\min} = 0$ , while in Fig. 3.5(b) the minimum rate guarantee is assumed to be 50 Kb/s, i.e.,  $x_{\min} = 50$  Kb/s. Procedure 2 was used to obtain the optimal EJT, and the tradeoff achieved by the  $\alpha$ -fair scheduler was obtained by solving (2.2). The convex optimization problems underlying Procedure 2 and the  $\alpha$ -fair scheduler were solved using ‘fmincon’, which is available in the MATLAB optimization toolbox. From both figures, it can be seen that, while these tradeoffs are close to each other for small and large values of  $\alpha$ , for intermediate values, the tradeoff generated by the optimal EJT scheduler is significantly better than that generated by the  $\alpha$ -fair one.



**Figure 3.5:** A comparison between the EJT's achieved by the optimal and  $\alpha$ -fair schedulers with time-sharing.

For example, for a Jain's index of 0.7, the optimal EJT scheduler provides 33% gain in efficiency as compared to the  $\alpha$ -fair scheduler for the case of  $x_{\min} = 0$  Kb/s and 21% gain in efficiency for the case of  $x_{\min} = 50$  Kb/s.

In Figs. 3.6(a) and 3.6(b) we present the counterparts of Figs. 3.5(a) and 3.5(b), but for the case when the RBs are not time-shared by the users. In Fig. 3.6(a),  $x_{\min} = 0$ , while in Fig. 3.6(b)  $x_{\min} = 50$  Kb/s. Since the set  $\mathcal{C}$  in this case is not convex and hence does not satisfy the sufficient conditions for possessing the monotonic tradeoff property, Procedure 1 was used to obtain the optimal EJT; cf. Section 3.3. The non-convex optimization problems underlying Procedure 1 and the  $\alpha$ -fair scheduler were solved using exhaustive search. Similar to the case of time-sharing considered in Figs. 3.5(a) and 3.5(b), it can be seen from Figs. 3.6(a) and 3.6(b) that, in this case too, the optimal EJT scheduler provides tradeoffs that are significantly better than those provided by the  $\alpha$ -fair scheduler when  $x_{\min} = 0$  and when  $x_{\min} = 50$  Kb/s.



**Figure 3.6:** A comparison between the EJT's achieved by the optimal and  $\alpha$ -fair schedulers without time-sharing.

### 3.5 Radio Resource Allocations in Ergodic Time-Varying Channels: Analysis

We now consider the problem of determining the radio resource allocations that span the optimal EJT when the channels are ergodic and time-varying. Before providing the mathematical framework for this case, we begin by noting that, from a practical perspective, one is typically interested in average, rather than instantaneous, rates [23,30]. In those cases, one might be tempted to apply the same approach in the previous section on instantaneous realizations of the channels. Although this would guarantee optimal tradeoff between efficiency and fairness in every time instant, it suffers from a major drawback that was alluded to in [23]. In particular, applying that instantaneous strategy does not necessarily lead to long-term average rates that are optimal from an EJT perspective.

For ergodic time-varying channels considered in this section, the channel gains assume random values in every sub-frame  $t$ . These gains are assumed to be available at the transmitter and the receivers, and the instantaneous data rate that can be achieved by each user  $m \in \mathcal{M}$  on each RB  $n \in \mathcal{N}$  is denoted by  $r_{m,n}(t)$ . To

achieve various points on the long-term average optimal EJT, we define scheduling variables, similar to Section 3.3. However, in the current case of time-varying channels, these variables are binary, indexed by  $t$ , and denoted by  $\{\rho_{m,n}(t)\}$ . The reason that  $\{\rho_{m,n}(t)\}$  are assumed to be binary is that the channel gains take on different values in each sub-frame, rendering time-sharing implausible. In other words, for each channel realization, the role of  $\{\rho_{m,n}(t)\}$  is to assign each RB to a particular user. Updating  $\{\rho_{m,n}(t)\}$  to achieve points on the long-term average optimal EJT will be accomplished using the gradient scheduling algorithm, which we describe next.

### 3.5.1 The Gradient Scheduling Algorithm

The gradient scheduling algorithm is a particular instance of adaptive algorithms that enable efficient solving of stochastic optimization problems in which the utilities to be maximized involve long-term averaging over an ergodic process; see e.g., [23,31]. The key idea that underlies such an algorithm is to use gradient-based steps to update the optimization variables sequentially using current and previous observations of the process. In addition to its relative simplicity, variants of the gradient scheduling algorithm were shown in [23] and [31] to yield the optimal solution of the stochastic optimization problem as the number of observations becomes sufficiently large.

To apply this algorithm to the current framework, let  $R_m(t)$  be the data rate scheduled to user  $m$  at sub-frame  $t$ , i.e.,  $R_m(t) = \sum_{n=1}^N \rho_{m,n}(t)r_{m,n}(t)$ . Maximizing the standard average rate of user  $m$  directly results in spurious behaviour [32], which can be alleviated by using the exponentially-weighted moving average instead. To do so, let  $\mu \in (0, 1)$  be a small positive scalar [31] and define  $W_m(t)$  to be

$$\begin{aligned}
W_m(t) &= \mu \sum_{i=0}^t (1 - \mu)^{i-t} R_m(i) \\
&= (1 - \mu)W_m(t - 1) + \mu R_m(t).^2
\end{aligned} \tag{3.25}$$

For notational convenience, let  $\mathbf{W}(t) = [W_1(t), \dots, W_M(t)]^T$  and  $\mathbf{R}(t) = [R_1(t), \dots, R_M(t)]^T$ . Since our goal is to optimize long-term (i.e., steady-state) average rates, the benefit vector of the  $M$  users can be defined to be  $\mathbf{x} = \lim_{t \rightarrow \infty} \mathbf{W}(t)$ .

Using the above notation, the idea behind gradient scheduling algorithm can be described as follows: Given the exponentially-weighted average rates at sub-frame  $t - 1$ ,  $\mathbf{W}(t - 1)$ , and the instantaneous rates,  $\{r_{m,n}(t)\}$ , the task of the scheduler is to determine the instantaneous scheduling variables,  $\{\rho_{m,n}(t)\}$ , in such a way that maximizes a given system utility  $U(\mathbf{x}) : \mathbb{R}^M \rightarrow \mathbb{R}$ . We will later show how  $U(\cdot)$  can be chosen to account for various tradeoff criteria. Since at sub-frame  $t$  the scheduler knows the previous values of  $\mathbf{W}(t)$ , but not future ones, its instantaneous decisions,  $\{\rho_{m,n}(t)\}$ , can only depend on  $\mathbf{W}(t - 1)$  and  $\{r_{m,n}(t)\}$ . In the gradient scheduling algorithm, the scheduler generates these decisions using the first order Taylor's series expansion of  $U(\mathbf{W}(t))$  around  $\mathbf{W}(t - 1)$ . In particular, using (3.25) with a sufficiently small  $\mu$ , we can write  $U(\mathbf{W}(t)) \approx U(\mathbf{W}(t - 1)) + \mu \nabla U(\mathbf{W}(t - 1))^T (\mathbf{R}(t) - \mathbf{W}(t - 1))$ . Noting that, at sub-frame  $t$ ,  $U(\mathbf{W}(t - 1))$  is constant, it can be seen that only the term containing  $\mathbf{R}(t)$  depends on  $\{\rho_{m,n}(t)\}$ . Hence, maximizing  $U(\mathbf{W}(t))$  is approximately equivalent to solving

$$\max_{\{\rho_{m,n}(t)\} \in \mathcal{S}} \nabla U(\mathbf{W}(t - 1))^T \mathbf{R}(t), \tag{3.26}$$

---

<sup>2</sup>It can be seen from (3.25) that the change in  $W_m(t)$  in every sub-frame  $t$  is given by  $W_m(t) - W_m(t - 1) = \mu(R_m(t) - W_m(t - 1))$ . Since  $R_m(t)$  is a random process,  $W_m(t)$  may oscillate and the amount of oscillation is proportional to  $\mu$ . Intuitively, to reduce (or possibly eliminate) the oscillation and improve the convergence of  $W_m(t)$ , it is necessarily to choose  $\mu$  to be sufficiently small, as we will discuss shortly.

where  $\mathcal{S} \triangleq \{\{\rho_{m,n}\} | \sum_{m=1}^M \rho_{m,n} \leq 1, \rho_{m,n} \in \{0, 1\}, \forall m \in \mathcal{M}, n \in \mathcal{N}\}$ .

Invoking the definition of  $\mathbf{R}(t)$ , the solution of (3.26) can be expressed as

$$\rho_{m,n}(t) = \begin{cases} 1, & \text{if } m = \arg \max_{m \in \mathcal{M}} \frac{\partial U(\mathbf{W}(t-1))}{\partial W_m(t-1)} r_{m,n}(t), \\ & \text{and } \max_{m \in \mathcal{M}} \frac{\partial U(\mathbf{W}(t-1))}{\partial W_m(t-1)} \geq 0, \\ 0, & \text{otherwise.} \end{cases} \quad (3.27)$$

It is shown in [23, 31, 32] that, when the rate processes  $\{r_{m,n}(t)\}$  are ergodic and the utility  $U(\cdot)$  is concave, the scheduling variables obtained by the gradient scheduling algorithm in (3.27) yield a long-term average rate vector  $\mathbf{x}$  that maximizes  $U(\mathbf{x})$ , asymptotically as  $\mu \rightarrow 0$ .

### 3.5.2 Application of Gradient Scheduling to Efficiency-Fairness Utilities

With a proper choice of  $U(\cdot)$  in (3.27), the gradient scheduling algorithm can be made to yield instantaneous schedules that attain long-term optimal efficiency-fairness tradeoffs for various fairness measures. To show this, we consider the case in which the set  $\mathcal{C}$  contains the long-term average rate benefit vectors corresponding to all possible choices of the scheduling variables  $\{\rho_{m,n}(t)\}$ , i.e.,

$$\mathcal{C} = \bigcup_{\{\rho_{m,n}(t)\} \in \mathcal{S}, \forall t} \left\{ \mathbf{x} \mid x_m = \lim_{t \rightarrow \infty} \mu \sum_{i=0}^t \sum_{n=1}^N (1 - \mu)^{i-t} \rho_{m,n}(i) r_{m,n}(i) \right\}. \quad (3.28)$$

This definition implies that the instantaneous scheduling variables generated by (3.27) yield long-term average rate benefit vectors that lie in  $\mathcal{C}$ .

### Achieving $\alpha$ -Fairness

When the fairness measure is given by the  $\alpha$ -fair utility, the function  $U(\cdot)$  in (3.27) is replaced with the utility  $U_\alpha(\mathbf{x})$  in (2.3). Since this utility is concave for all  $\alpha \in [0, \infty)$ , the gradient scheduling algorithm with  $\mu \rightarrow 0$  can be used to obtain the instantaneous schedules that yield the corresponding long-term optimal average rate vectors. In this case, these schedules are given by

$$\rho_{m,n}(t) = \begin{cases} 1, & \text{if } m = \arg \max_{m \in \mathcal{M}} (W_m(t-1))^{-\alpha} r_{m,n}(t), \\ 0, & \text{otherwise.} \end{cases} \quad (3.29)$$

### Achieving the Optimal EJT

We now show how to use the gradient scheduling algorithm to obtain the optimal EJT. For simplicity we restrict our attention to the case of  $x_{\min} = 0$ . To consider this case, we note that, with  $x_{\min} = 0$ , the set of feasible benefit vectors defined in (3.28) satisfies the conditions of Theorem 1. In particular, this set is convex and contains the all-zero vector,  $\mathbf{0}_M$ . To see that  $\mathcal{C}$  is convex in  $\mathbf{x}$ , we note that, for any two long-term average rate benefit vectors  $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{C}$  and any  $\theta \in [0, 1]$ , the line segment  $\theta \mathbf{x}_1 + (1 - \theta) \mathbf{x}_2$  is also in  $\mathcal{C}$  [23]. That  $\mathbf{0}_M \in \mathcal{C}$  follows from the fact that setting the scheduling variables  $\rho_{m,n}(t) = 0$  for all  $m, n$  and  $t$  is feasible, i.e., the all-zero  $M \times N$  matrix,  $\mathbf{0}_{MN} \in \mathcal{S}$ .

Now that the conditions of Theorem 1 are satisfied, we know that  $\mathcal{C}$  possesses the monotonic tradeoff property in Definition 3, and Procedure 2 can be used to find the optimal EJT long-term average rates. In this procedure, the optimization problem in (2.21) is solved for various choices of  $\sigma$ . For each value of  $\sigma$ , the problem in (2.21) involves a constraint on the sum of the long-term average rates. Unfortunately, incorporating such a constraint in the gradient scheduling algorithm is not

straightforward and hence, this algorithm cannot be used directly to solve (2.21) in the current stochastic framework. To circumvent this difficulty, we use the alternate formulation of (2.21) given in (2.26). Using that formulation, the gradient scheduling algorithm can be applied with the utility  $U(\cdot)$  in (3.27) replaced with

$$U_\beta(\mathbf{x}) = -\|\mathbf{x}\|^2 + 2\beta\eta(\mathbf{x}), \quad (3.30)$$

where, as explained in Section 2.4.5,  $\beta \in [0, \infty)$ .

Using the utility in (3.30), the instantaneous schedules that yield the optimal EJT long-term average rate vectors, for a given  $\beta$ , are given by

$$\rho_{m,n}(t) = \begin{cases} 1, & \text{if } m = \arg \max_{m \in \mathcal{M}} (\beta - W_m(t-1))r_{m,n}(t), \\ & \text{and } \max_{m \in \mathcal{M}} (\beta - W_m(t-1)) \geq 0, \\ 0, & \text{otherwise.} \end{cases} \quad (3.31)$$

### Achieving $\alpha$ -Fairness and Optimum EJT with Minimum Rate Guarantee

To ensure that the long-term average rate of each user exceeds a given threshold,  $x_{\min}$ , both the  $\alpha$ -fair and the optimal EJT schedulers can be modified to incorporate the token-based procedure developed in [33]. This procedure can be used to maximize any concave utility in the long-term average rate while satisfying minimum average rates constraints. By applying this procedure to maximize the  $\alpha$ -fair utility given in (2.3) and to maximize the optimal EJT utility given in (3.30), we get the following  $\alpha$ -fair and optimal EJT schedulers, with minimum rate guarantee:

$$\rho_{m,n}(t) = \begin{cases} 1, & \text{if } m = \arg \max_{m \in \mathcal{M}} \{e^{\alpha\kappa_m(t)}(\beta - W_m(t-1))r_{m,n}(t)\}, \\ 0, & \text{otherwise,} \end{cases} \quad (3.32)$$

$$\rho_{m,n}(t) = \begin{cases} 1, & \text{if } m = \arg \max_{m \in \mathcal{M}} \{e^{a\kappa_m(t)} (W_m(t-1))^{-\alpha} r_{m,n}(t)\}, \\ & \text{and } \max_{m \in \mathcal{M}} (\beta - W_m(t-1)) \geq 0, \\ 0, & \text{otherwise,} \end{cases} \quad (3.33)$$

where  $a$  is a multiplicative step-size that must be properly adjusted and  $\kappa_m(t)$  is updated as follows

$$\kappa_m(t) = \max\{0, \kappa_m(t-1) + x_{\min} - R_m(t-1)\}, \quad (3.34)$$

where  $\kappa_m(0) = 0$  for all  $m \in \mathcal{M}$ , and  $R_m(t)$  is the instantaneous data rate scheduled to user  $m$  at sub-frame  $t$ . Intuitively,  $\kappa_m(t-1)$  plays the role of increasing the weight of the users that did not meet the rate requirement at sub-frame  $t-1$  and decreases the weight of the users that exceeds the requirement at sub-frame  $t-1$ . Optimality proof and further details about this procedure can be found in [33].

We note, however, that this procedure requires intricate tuning of several operation parameters, including a multiplicative step-size  $a$  and a practical operation margin on  $x_{\min}$ . Hence, although this technique offers a means for enforcing a minimum rate requirement, it does that indirectly and its effectiveness depends, to a large extent, on the precision with which the operation parameters are tuned. As indicated in [33], if the user long-term average rates converge, then they converge to the optimal long-term average rates that maximize the desired utility while meeting the minimum rate guarantee; however, convergence itself is not guaranteed and it depends on parameter tuning.

An important special case of the scheduler given by (3.33) is the maximum-throughput with minimum-rate (MTMR) scheduler proposed in [33], which corresponds to the case of  $\alpha = 0$ . The goal of the MTMR scheduler is to maximize the sum-rate efficiency while satisfying the minimum rate requirement,  $x_{\min}$ , and it is

given by

$$\rho_{m,n}(t) = \begin{cases} 1, & \text{if } m = \arg \max_{m \in \mathcal{M}} \{e^{a\kappa_m(t)} r_{m,n}(t)\}, \\ 0, & \text{otherwise.} \end{cases} \quad (3.35)$$

The scheduler given by (3.35) is parameterized by  $x_{\min}$  and it can also be used in trading off efficiency with fairness by varying the parameter  $x_{\min}$ .

Now that we have shown how the gradient scheduling algorithm can be used to yield  $\alpha$ -fair and optimal EJT long-term average rate vectors, in the next section we will investigate the performance of this algorithm in practical wireless communication scenarios.

## 3.6 Radio Resource Allocations in Ergodic Time-Varying Channels: Simulation Results

We now compare the performance of the  $\alpha$ -fair scheduler in (3.29) and the MTMR scheduler in (3.35) with the proposed optimal EJT in (3.31), in ergodic time-varying channels. To do so, we start by providing the simulation setup and parameters in Section 3.6.1. We then discuss the steps taken to validate the accuracy of our simulator in Section 3.6.2, and we conclude this section by providing comparison results in Section 3.6.3.

### 3.6.1 Simulation Setup and Parameters

We consider a cellular network based on the International Mobile Telecommunications-Advanced (IMT-Advanced) guidelines for the Urban Macro (UMa) scenario [34]. As per these guidelines, a hexagonal layout with wraparound is considered with 57 hexagonal sectors and 10 UTs per sector. These sectors are served by 19 BSs, each

with a tri-sector antenna to serve a 3-sector cell-site. The users are uniformly dropped in the 57 sectors, and Monte Carlo simulations are carried over  $10^4$  sub-frames and averaged over 10 independent drops. The users are assumed to be associated with the sector with the highest received average power. This power depends on large channel variations, which are mainly due to non-uniform antenna patterns (cf. [34, pp. 17–18]) and distance-dependent path-loss and correlated shadowing. To incorporate these variations in our simulations, we considered the line-of-sight (LOS) and non-line of-sight (NLOS) users as in [34, pp. 17–33]. For the LOS users, the path-loss exponent is set to be 2.2 and the shadowing is assumed log-normal with standard deviation of 4 dB. For the NLOS users, the path-loss exponent is set to be 3.9 and the log-normal shadowing standard deviation is set to be 6 dB. The system parameters are based on the LTE standard, whereby each RB is composed of a time slot of 7 OFDM symbols and 12 subcarriers [29]. The simulation parameters are given in Table 3.1.

In addition to large variations, the received signal power is subject to small-scale variations due to time-varying multipath fading. Variations of the received signal power due to fading within each RB is negligible and hence, the channels can be assumed fixed over each RB. For other RBs the channels take on different values depending on the spectro-temporal correlation of the IMT-Advanced model for the UMa scenario [35].

At sub-frame  $t$ , each user  $m \in \mathcal{M}$  calculates its received signal-to-interference-plus-noise ratios (SINRs) on all the RBs, and subsequently determines  $r_{m,n}(t)$ , the data rate that can be reliably communicated on every RB  $n \in \mathcal{N}$  using the adaptive modulation and coding (AMC) scheme given in Table 3.2<sup>3</sup>. This AMC table was devised using the AMC strategy given in [36]. The set of all the rates at sub-frame

---

<sup>3</sup>In LTE, a sub-frame spans 1 ms and consists of two time slots, i.e., there are  $2N$  RBs to be scheduled per sub-frame; however, based on the current simulation parameters, the channel stays the same for both time slots in the sub-frame (coherence time is 3 ms). Thus, it suffices to consider only  $N$  RBs for determining the scheduling decisions.

$t$ ,  $\{r_{m,n}(t)\}$  are transmitted through feedback control channels to the BS, which subsequently determines the appropriate scheduling variables,  $\{\rho_{m,n}(t)\}$ , depending on previously scheduled rates and the adopted scheduling strategy. Finally, we assume a reuse-1 scheme in which all radio resources are used in every sector<sup>4</sup>.

### 3.6.2 Validation of the Simulator

#### Wrap Around Implementation and Validation

Consider a network that consists of 57 hexagonal sectors as shown in Fig. 3.7. Clearly, the interior sectors (sectors 1 to 3) are subject to higher interference than the sectors at the boundary of the layout. An important metric to assess the effect of interference is wideband SINR, which is defined as the ratio of the average power received from the serving sector to the sum of the average power received from all other sectors and the noise power at the UT (i.e., small scale fading is not included in the calculation of the wideband SINR). To quantify the difference in interference levels between the interior and the boundary sectors, we plot in Fig. 3.8 the wideband SINR cumulative distribution function (CDF) for the interior and the boundary sectors, as well as the wideband SINR CDF of all sectors. This figure shows that the boundary sectors have significantly higher wideband SINR than the interior sectors; e.g., the median wideband SINR for the boundary sectors is 4 dB higher than that of the interior sectors. Since the wideband SINR CDF varies significantly across the sectors, it is not meaningful to collect user rates from all sectors. A trivial solution to this boundary-effect is to collect the rates from the users in the interior sectors only and discard the rates from all other users. Unfortunately, such a solution is not computationally-efficient as it discards  $\frac{54}{57}$  of the calculated rates.

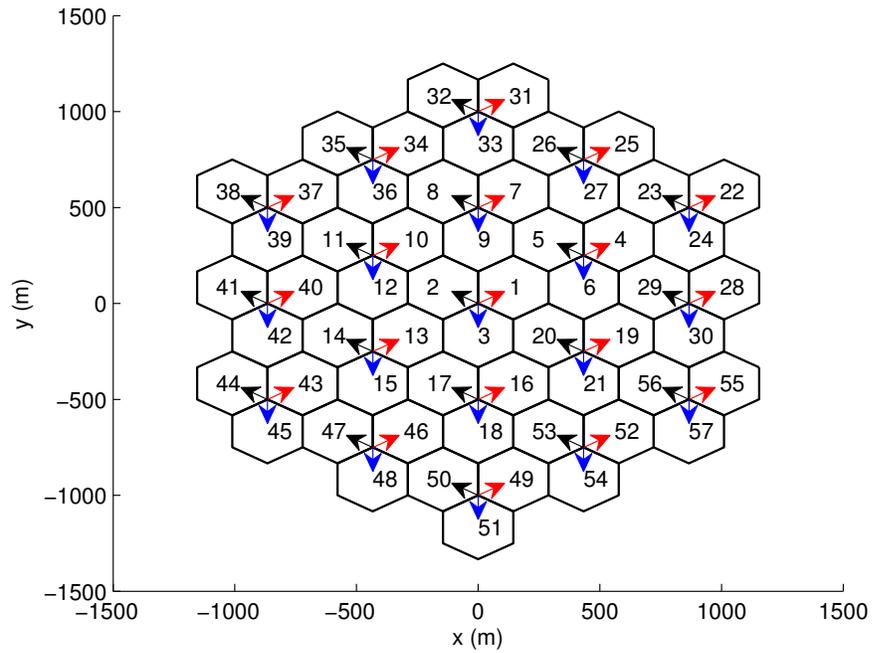
A computationally-efficient solution to the boundary-effect problem is to use a

---

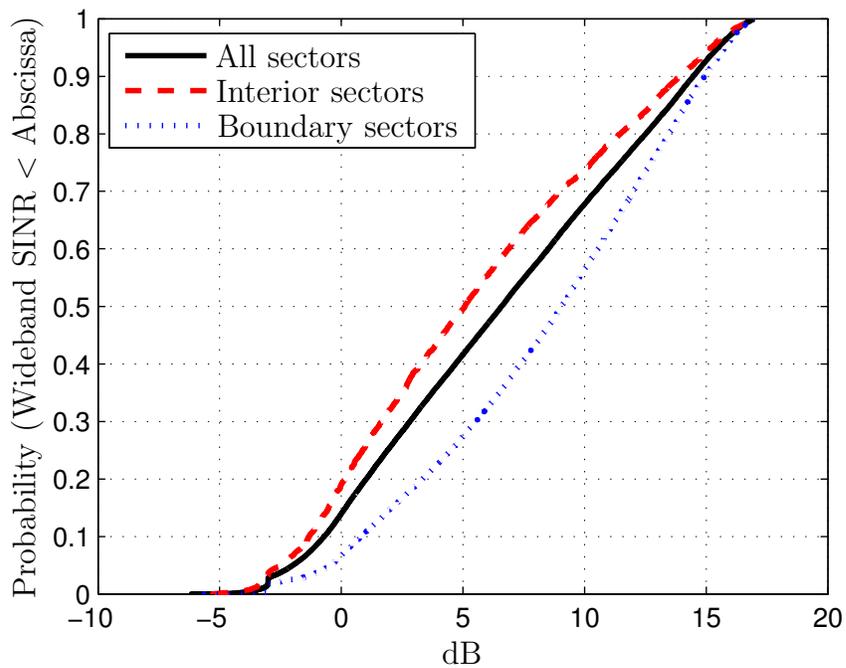
<sup>4</sup>The implications of relaxing this assumption to manage interference is investigated in depth in Chapter 4.

**Table 3.1:** Simulation parameters based on IMT-Advanced UMa scenario.

Parameter	Assumption
Number of sectors	57 (wraparound)
Number of UTs per sector	10
Inter-site distance	500 m
BS's height	25 m
Min. dist. between a UT and a BS	25 m
UT's speed	30 km/h
Bandwidth (downlink)	10 MHz
Sub-carrier spacing	15 KHz
Number of RBs ( $N$ )	50
OFDM symbol duration	66.67 $\mu$ s
Number of sub-carriers per RB	12
Number of OFDM symbols per RB	7
Number of drops	10
Number of sub-frames per drop	10000
Noise power per RB ( $P_n$ )	-114.45 dBm
Carrier frequency	2.0 GHz
Total BS's transmit power	46 dBm
Path loss and Shadowing	Based on UMa scenario [34]
Smoothing factor ( $\mu$ )	0.01
BS antenna gain (boresight)	17 dBi
User antenna gain	0 dBi
Feeder loss	2 dB
Channel estimation delay	4 ms
SINR estimation margin (to reduce the effect of outdated CSI)	6 dB
BS's antenna tilt ( $\phi_t$ )	12° [37]
BS's horizontal antenna pattern	$A(\theta) = -\min [12(\frac{\theta}{70^\circ})^2, 20 \text{ dB}]$ [34]
BS's elevation antenna pattern	$A_e(\phi) = -\min [12(\frac{\phi-\phi_t}{15^\circ})^2, 20 \text{ dB}]$ [34]
BS's combined antenna pattern	$-\min [- (A(\theta) + A_e(\phi)), 20 \text{ dB}]$ [34]
Small-scale fading model	IMT-Advanced channel model [35]
Traffic model	Full buffer



**Figure 3.7:** A network layout that consists of 57 hexagonal sectors (without wrap around).



**Figure 3.8:** CDF of the wideband SINR for the network layout shown in Fig. 3.7.

**Table 3.2:** The AMC table which is used to determine the achievable data rate on each RB for each user based on the SINR. This table is based on the AMC strategy given in [36].

SINR range on RB $n$ for user $m$ (dB)	Data rate, $r_{m,n}$ , (kb/s)
$(-\infty, -6.1]$	0
$(-6.1, -4.1]$	35.3
$(-4.1, -2.0]$	56.4
$(-2.0, -0.2]$	92.4
$(-0.2, 1.9]$	131.4
$(1.9, 3.8]$	177.4
$(3.8, 5.8]$	223.1
$(5.8, 8.5]$	291.6
$(8.5, 9.9]$	388.4
$(9.5, 12.5]$	418.3
$(12.5, 14.8]$	544.3
$(14.8, 16.1]$	648.1
$(16.1, 17.8]$	721.7
$(17.8, \infty)$	807.4

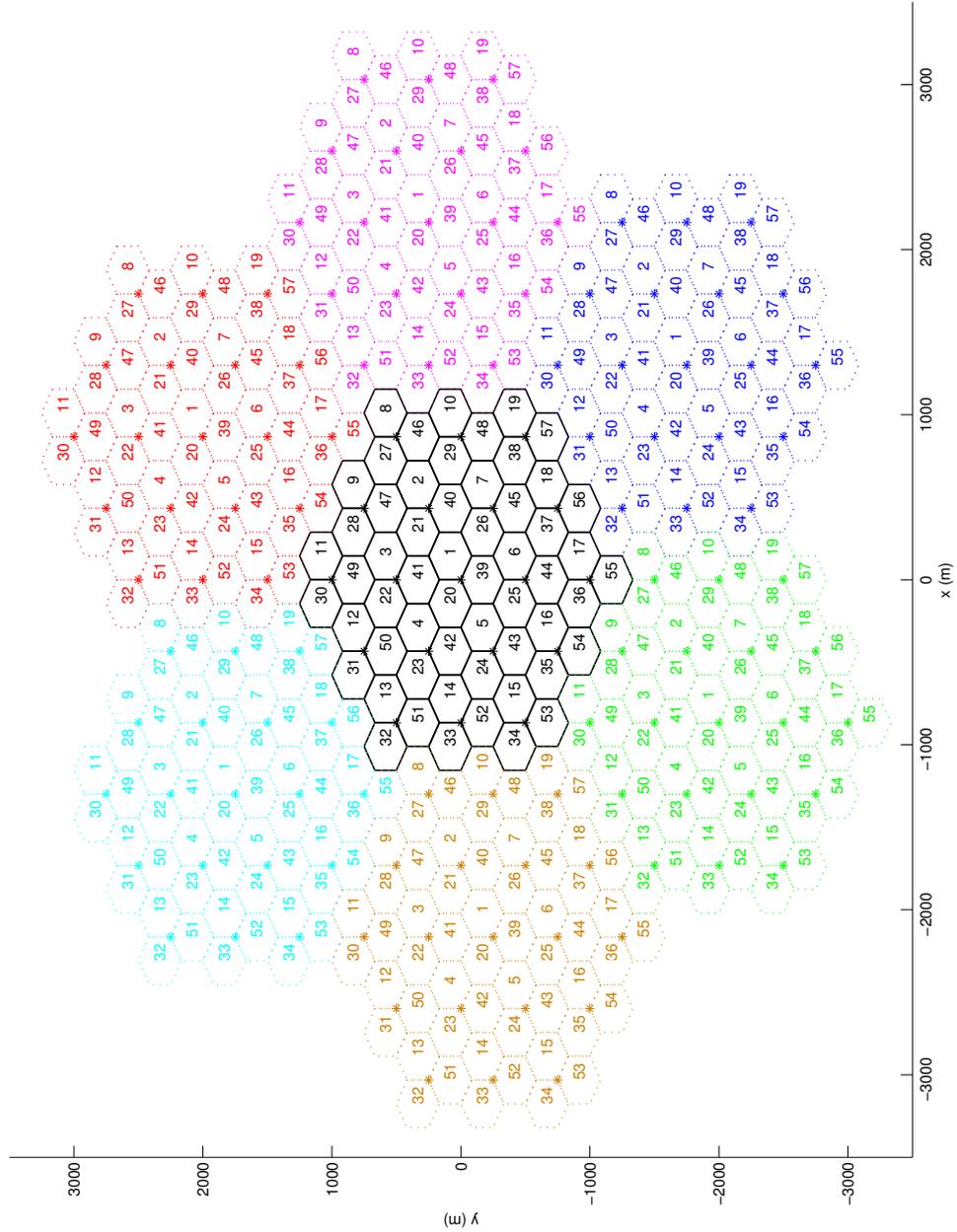
wrap around layout [38, Appendix G]. In a nutshell, the boundary-effect can be significantly reduced by adding “virtual” 6 clusters around the original cluster of 57 hexagonal sectors, as shown in Fig. 3.9. As a result, both the interior and the boundary sectors in the original cluster have the same number of neighboring interferers. The details of constructing and using wrap around layout can be found in [38, Appendix G]. To examine the boundary-effect on the wrap around layout, we plot in Fig. 3.10 the wideband SINR CDF for the interior and the boundary sectors, as well as the wideband SINR CDF of all 57 sectors (sectors in the “virtual” 6 clusters are ignored). It can be seen from this figure that the interior, the boundary, and all the sectors have the same wideband SINR CDF, which means that the boundary-effect

that we observed in Fig. 3.8 is illuminated. The agreement in the wideband SINR CDF also validates the implementation of the wrap around layout. Hence, we can now collect the rates from users in all 57 sectors; this results in a significant reduction in the simulation time as compared to collecting the rates from the interior sectors only.

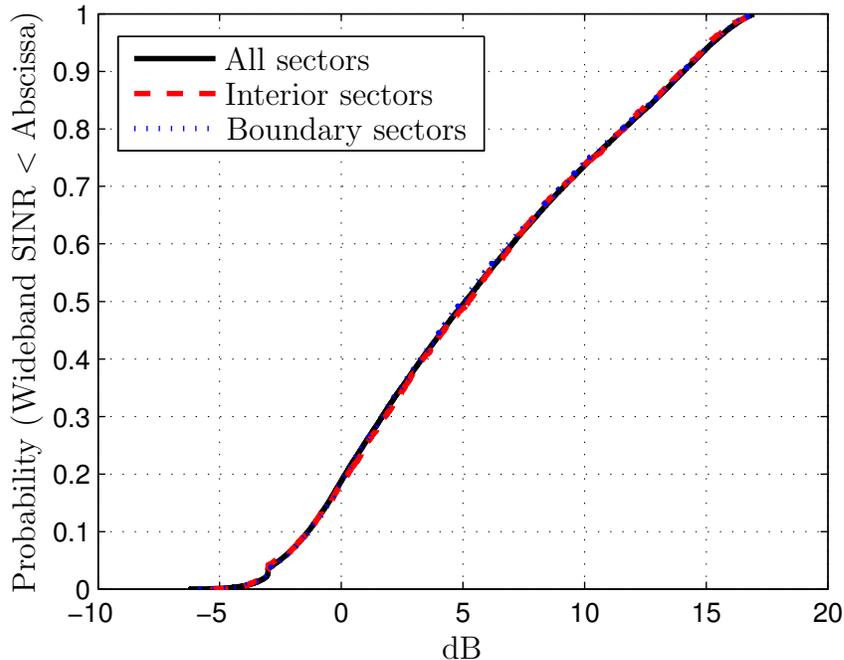
### Wideband SINR CDF Validation

System simulators are complex in nature which make them error-prone. As a result, careful verification of the simulation results is imperative. Consequently, we compare the results obtained by the MATLAB-based simulator we developed with the ones produced by WINNER+ research partners using different simulation tools for UMa scenario [37].

We start by highlighting an aspect in the simulation that has important effect on assessing the performance of cellular systems. In particular, we demonstrate that the way UTs are associated to sectors has a significant impact on the simulation results. In the following, we explain three association strategies, namely, wideband SINR-based, wideband SINR-based (excluding shadowing), and geographical-based association. In wideband SINR-based association, each UT is associated with the antenna sector to which it has the highest wideband SINR. This association strategy resembles reality and provides the most favorable results in terms of wideband SINR. A consequence of this strategy is that the coverage region of each sector is not hexagonal and it changes from drop to drop due to the different shadowing realizations. Another consequence is that a UT may not be associated to the closest sector antenna as it may experience heavy shadowing to that sector antenna. A more convenient way of doing simulation is to exclude shadowing in the calculation of the wideband SINR which leads to fixed coverage region for each sector in all drops. Due to the directional antenna patterns, the coverage regions of each sector is not hexagonal. In geographical-based



**Figure 3.9:** A wrap around network layout that consists of a cluster of 57 hexagonal sectors (the middle cluster) and 6 “virtual” clusters.



**Figure 3.10:** CDF of the wideband SINR for the network layout shown in Fig. 3.9.

association, a UT is associated to a particular sector antenna if it resides inside the hexagonal area of that sector. In this strategy, the coverage region of each sector is hexagonal.

In Fig. 3.11, we plot the CDF of the wideband SINR using the three association strategies. For calibration purposes, we also include a reference CDF that is obtained by averaging the CDF results produced by seven WINNER+ partners using different simulation tools [37]. It is clear from the figure that wideband SINR-based association strategy produces a CDF that agree very well with the reference results; however, the other two association strategies produce a heavy tail which would impact the throughput of the users at the cell edge. The heavy tail is a direct consequence of associating UTs to sectors in a suboptimal manner. As a result, these two association strategies may not be suitable for assessing the performance, especially for the interference-aware schemes that we will consider in Chapter 4, as these association

strategies tend to exaggerate the gains achieved for UTs at the cell-edge. This is the case because the cell-edge user throughput for these two association strategies is very low. This makes any improvement in cell-edge user throughput to be large as compared to the very low values of reuse-1. Consequently, we use wideband SINR-based association strategy throughout the thesis.

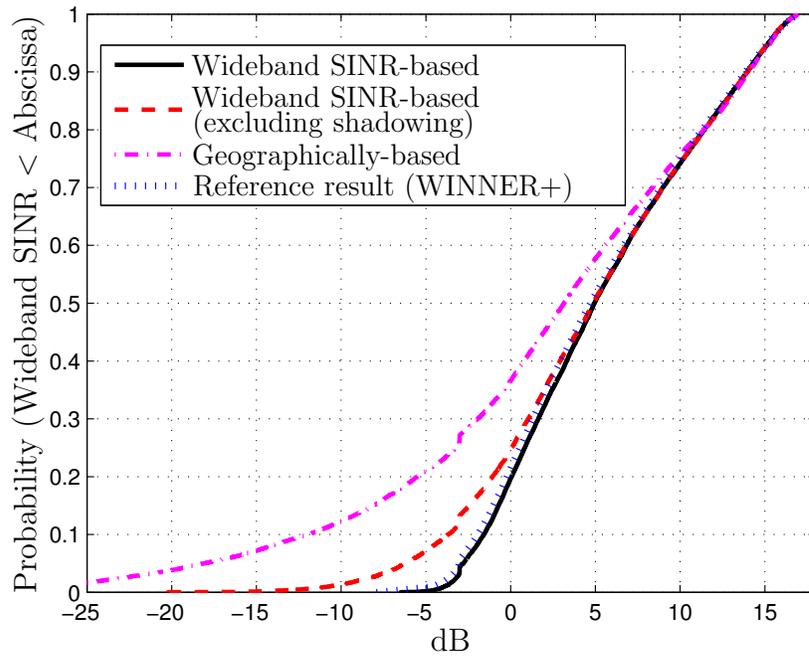
### **Path Gain CDF Validation**

In addition to checking the CDF of the wideband SINR, we also plot in Fig. 3.12 the distribution of the path gain, which is defined as the average difference between received and transmitted power in dB. In this figure, error bars represent the range of values obtained by seven WINNER+ partners [37]. It is clear that the distribution produced by our simulator is within the range of values obtained by WINNER+ partners.

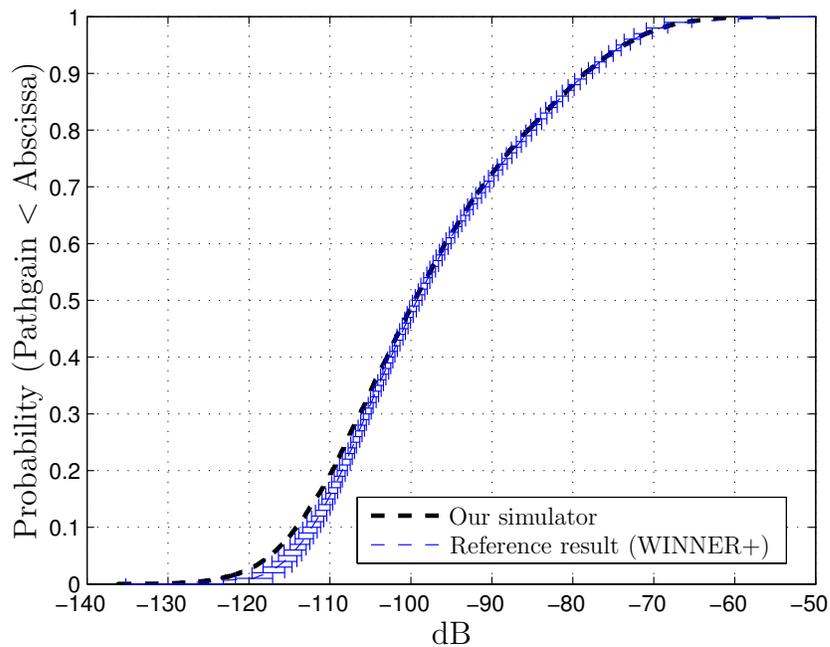
### **Throughput CDF Validation**

Finally, we plot the distribution of the normalized time-average user throughput using our simulator in Fig. 3.13 assuming a proportional-fair scheduler. Normalization is performed by dividing the user throughput over the total downlink bandwidth, which is 10 MHz. For the sake of comparison, we also present the throughput CDFs obtained by WINNER+ partners in Fig. 3.14 (reprinted from [37, Fig. 23]). There is slight discrepancy between the curves obtained by different industrial partners in WINNER+. The CDF produced by our simulator is within the range of CDFs obtained by WINNER+ partners.

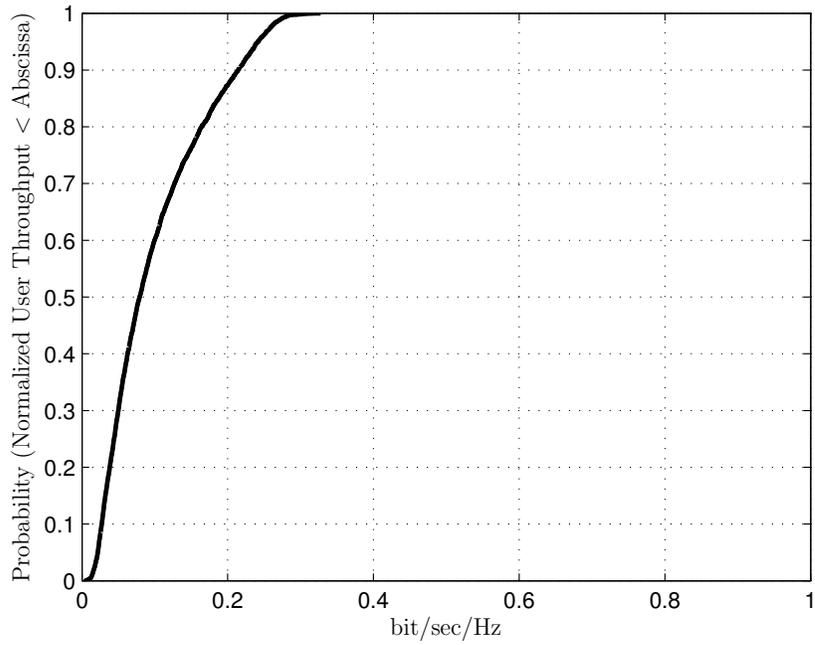
After validating the CDF of wideband SINR, CDF of pathgain, and CDF of average user throughput, we are ready to present the performance results in the next section.



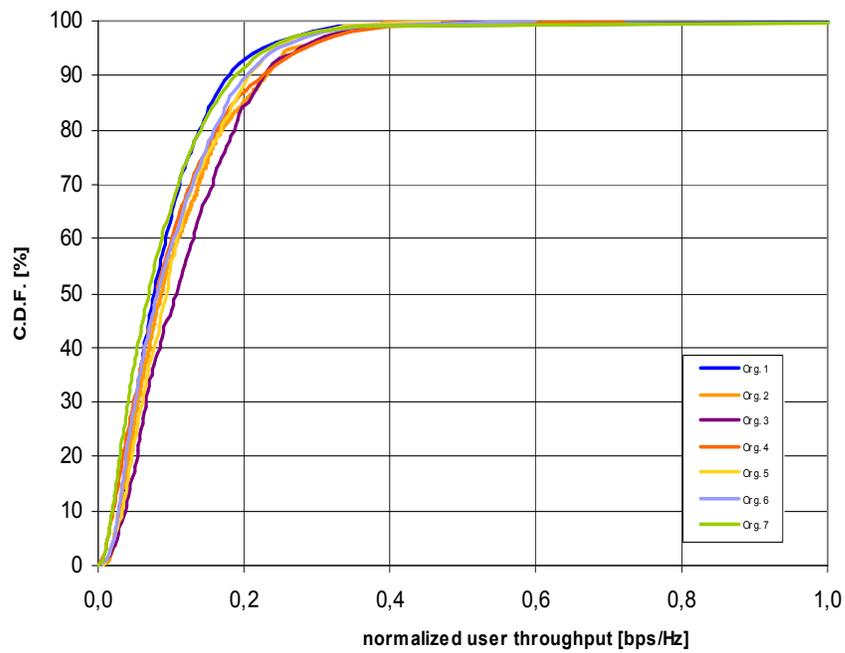
**Figure 3.11:** CDF of the wideband SINR for different association strategies. Reference result (WINNER+) refers to a reference CDF that is obtained by averaging the CDF results produced by seven WINNER+ partners [37].



**Figure 3.12:** CDF of the pathgain. Reference result (WINNER+) refers to a reference CDF that is obtained by averaging the CDF results produced by seven WINNER+ partners [37].



**Figure 3.13:** Distribution of the normalized time-average user throughput obtained using our simulator.



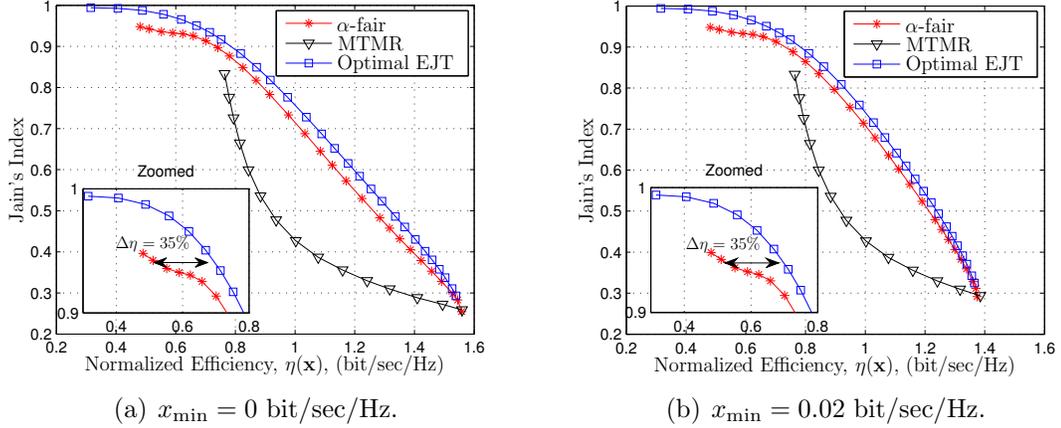
**Figure 3.14:** CDF of the normalized time-average user throughput produced by seven WINNER+ partners (reprinted from [37, Fig. 23]).

### 3.6.3 Comparison Results

To be consistent with the IMT-Advanced guidelines, we present the comparison results in terms of normalized throughputs measured in (bit/sec/Hz), where the term “normalized throughput” is defined as the long-term average rate divided by the total bandwidth (10 MHz). Moreover, normalized efficiency is defined as the normalized aggregate sector throughput which is obtained by summing the normalized user throughput of all users in the network and dividing the result by the number of sectors.

#### Efficiency-Jain Tradeoff Comparison

We plot in Fig. 3.15(a) the EJT curves assuming there is no minimum normalized throughput guarantee, for the  $\alpha$ -fair scheduler in (3.29), the MTMR scheduler in (3.35), and the proposed optimal EJT scheduler in (3.31). Similarly, we plot in Fig. 3.15(b) the EJT curves assuming the minimum normalized throughput guarantee to be  $x_{\min} = 0.02$  bit/sec/Hz, for the modified  $\alpha$ -fair scheduler in (3.33), the MTMR scheduler in (3.35), and the modified optimal EJT scheduler in (3.32). From both figures, it can be seen that the proposed optimal EJT scheduler achieves consistently better tradeoffs, especially in the high fairness region. For example, for a Jain’s index of 0.94, the proposed optimal EJT scheduler provides 35% gain in efficiency as compared to the  $\alpha$ -fair scheduler. Moreover, the MTMR scheduler produces the least favorable EJT as expected, since it accounts only for minimum-throughput users and not necessarily accounts for fairness among all users. In addition, we also observe that adding minimum throughput guarantee reduces the gap between the  $\alpha$ -fair scheduler and the optimal EJT one in the high efficiency region.



**Figure 3.15:** A long-term average rates comparison between the EJT's achieved by the  $\alpha$ -fair, the MTMR, and the proposed optimal EJT schedulers.

### User Satisfaction Comparison

To illustrate the advantage of using Jain's index as a fairness measure, we compare the number of satisfied users when the scheduling is performed using the optimal EJT,  $\alpha$ -fair, and the MTMR schedulers. In particular, we assume that a user is satisfied if its normalized throughput exceeds a certain threshold. Generally, we are interested in maximizing the number of satisfied users. A scheduler that achieves such maximization would be one that solves the following cardinality maximization problem:

$$\underset{\mathbf{x} \in \mathcal{C}}{\text{maximize}} \quad |\chi| \tag{3.36}$$

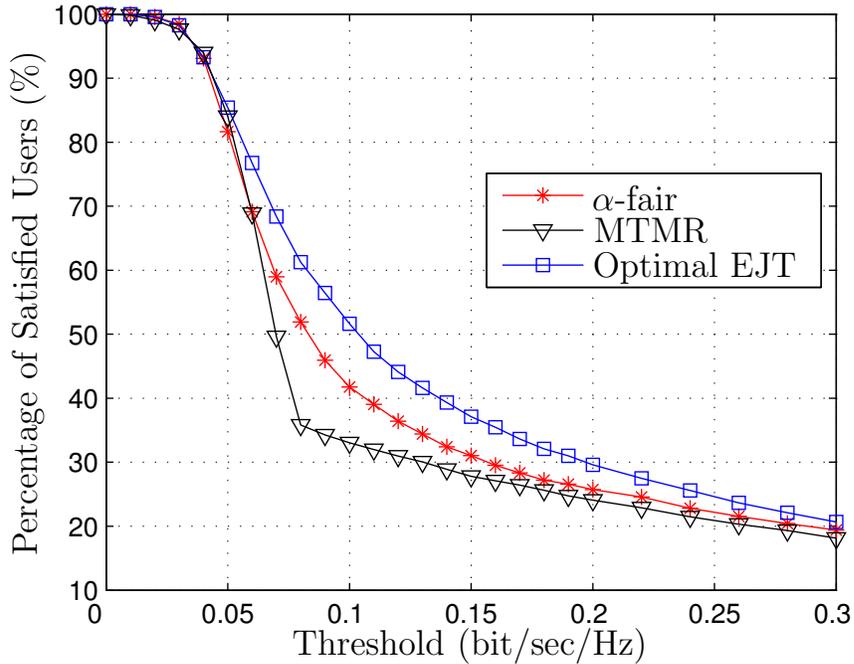
$$\text{subject to} \quad \chi = \{x_m \geq x_{\text{th}}, \forall m \in \mathcal{M}\}, \tag{3.37}$$

where  $x_{\text{th}}$  is the satisfaction threshold and  $|\chi|$  is the cardinality of the set  $\chi$  which is defined as the number of elements in  $\chi$ . Unfortunately, solving (3.36) is difficult and might incur NP-hardness, depending on the feasible set  $\mathcal{C}$ . One way, to alleviate this difficulty is to maximize  $|\chi|$  indirectly using an auxiliary scheduling objective. A natural candidate of such schedulers would be one that enables the allocation of resources

in a way that combines both efficiency and fairness. In our work, we have studied three such schedulers, namely, the  $\alpha$ -fair, the MTMR, and the proposed optimal EJT schedulers. Seeing as the objective of the optimal EJT scheduler is effectively the projection of the all-equal throughput vector on  $\mathcal{C}$ , it can be suspected that, in comparison with the  $\alpha$ -fair scheduler and MTMR, it enables a larger percentage of users to be satisfied.

To perform a concrete comparison, for the  $\alpha$ -fair, the MTMR, and the optimal EJT schedulers, the parameters  $\alpha$ ,  $x_{\min}$ , and  $\beta$ , respectively, are chosen to maximize the percentage of users for which the rates obtained by these schedulers exceed a given threshold. Unfortunately, obtaining closed-form expressions for the optimal  $\alpha$ ,  $\beta$ , and  $x_{\min}$  as functions of rate thresholds has proved intractable. As an alternative, for each threshold we performed an exhaustive search to find the optimal  $\alpha$ ,  $\beta$ , and  $x_{\min}$  that maximize the number of satisfied users in each scheduler. In practice, the optimal  $\alpha$ ,  $\beta$ , and  $x_{\min}$  can be determined through a pre-communication learning phase that enables the BSs to evaluate user satisfactions for tentative values of  $\alpha$ ,  $\beta$ , and  $x_{\min}$ .

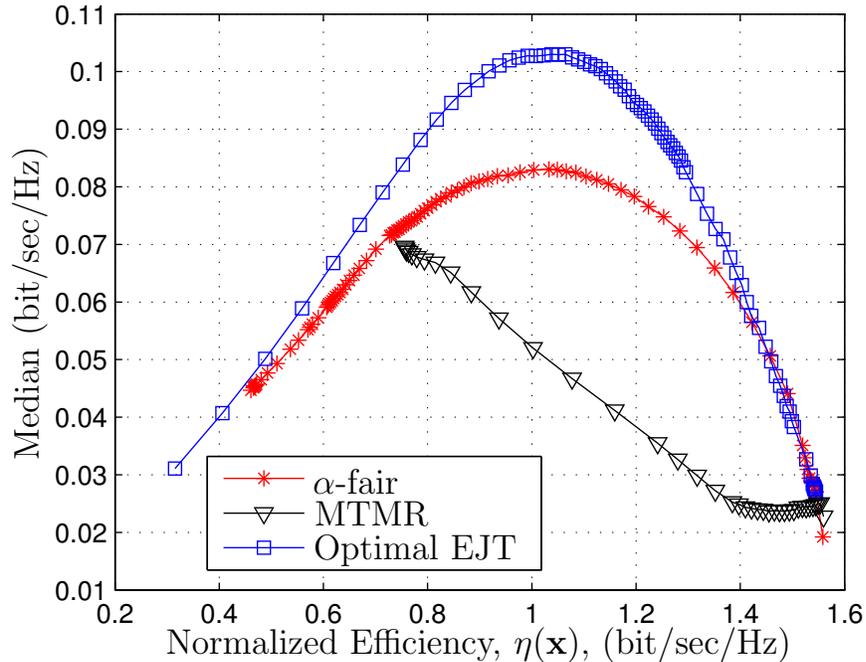
The comparison between the percentage of satisfied users at given normalized throughput thresholds is plotted in Fig. 3.16. As can be expected, the number of satisfied users is monotonically decreasing with the threshold for all schedulers. However, the optimal EJT scheduler is consistently better than the  $\alpha$ -fair and the MTMR schedulers as it enables significantly more users to be satisfied. For instance, for the considered scenario when the threshold is set to be 0.1 bit/sec/Hz, the scheduler based on the optimal EJT satisfies 10% and 18% more users than those satisfied by the  $\alpha$ -fair and the MTMR schedulers, respectively. Hence, from a service provider's perspective, using the scheduler based on the optimal EJT yields a valuable increase in the ability of the system to satisfy users with throughput requirements.



**Figure 3.16:** Percentage of satisfied users for different thresholds.

### Efficiency-Median Tradeoff Comparison

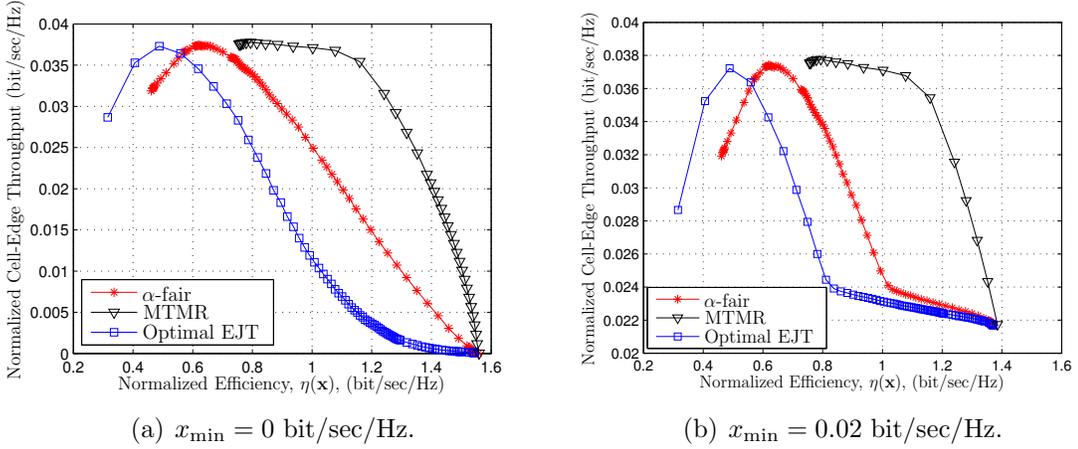
In addition to Jain’s index, an important metric that is related to fairness among users is the median, which is defined as the 50<sup>th</sup> percentile of the normalized throughput distribution (see e.g., [39]). In Fig. 3.17, we plot the Efficiency-Median Tradeoff curve for the  $\alpha$ -fair, the MTMR, and the proposed optimal EJT scheduler. It can be seen from this figure that the Efficiency-Median Tradeoff achieved by the optimal EJT scheduler is significantly better than those provided by the  $\alpha$ -fair and the MTMR schedulers. For example, for an efficiency of 1 bit/sec/Hz, the optimal EJT scheduler provides 24% and 97% gains in the median as compared to the  $\alpha$ -fair and the MTMR schedulers, respectively. Similar to the EJT curves presented in Fig. 3.15, the MTMR scheduler produces the least favorable Efficiency-Median Tradeoff since it accounts only for minimum-throughput users and not necessarily accounts for fairness among all users.



**Figure 3.17:** A long-term average rates comparison of the Efficiency-Median Trade-off achieved by the  $\alpha$ -fair, the MTMR, and the proposed optimal EJT schedulers.

### Efficiency-Cell-edge Tradeoff Comparison

We conclude the performance results in this section by presenting the Efficiency-Cell-edge Tradeoff curves for the  $\alpha$ -fair, the MTMR, and the proposed optimal EJT schedulers in Fig. 3.18(a) and Fig. 3.18(b) for  $x_{\min} = 0$  and  $x_{\min} = 0.02$  bit/sec/Hz, respectively. Normalized cell-edge user throughput is defined as the 5<sup>th</sup> percentile of the normalized throughput distribution. It can be seen from Fig. 3.18(a) that the MTMR scheduler provides consistently the best tradeoff. For instance, for a cell-edge normalized throughput of 0.02 bit/sec/Hz, the MTMR scheduler provides 28% and 62% gains in efficiency as compared to the  $\alpha$ -fair and the proposed optimal EJT schedulers, respectively. However, these large gains achieved by the MTMR scheduler come at the expense of reduced Jain's index, percentage of satisfied users, and the median normalized throughput, as demonstrated in Figs 3.15, 3.16, and 3.17,



**Figure 3.18:** A long-term average rates comparison of the Efficiency-Cell-edge Tradeoff achieved by the  $\alpha$ -fair, the MTMR, and the proposed optimal EJT schedulers.

respectively. Moreover, the behaviour of the optimal EJT scheduler for cell-edge users in time-varying channels agrees with the insight drawn from the closed-form expression of the optimal EJT rate vectors in quasi-static frequency-flat channels given by (3.21).

One way to improve the cell-edge performance of the  $\alpha$ -fair and the optimal EJT schedulers is to impose a minimum throughput guarantee that is strictly greater than 0, as shown in Fig. 3.18(b). As it can be seen from this figure, all three schedulers achieve cell-edge throughputs that is strictly greater than 0.02 bit/sec/Hz, as expected. While the MTMR still outperforms both the  $\alpha$ -fair and optimal EJT schedulers in the low efficiency region, the gap between the tradeoff curve achieved by the MTMR and the ones achieved by the  $\alpha$ -fair and the optimal EJT schedulers are significantly reduced in the high efficiency region. Hence, it is imperative to impose a minimum throughput guarantee that is strictly greater than 0 for both the  $\alpha$ -fair and the optimal EJT schedulers to strike a balance between efficiency, Jain's index, and cell-edge throughput.

### 3.7 Conclusions

In this chapter, we demonstrated that the conditions in Theorem 1 are naturally satisfied in various resource allocation problems in communication networks. In such problems, allocations that achieve the optimal EJT can be efficiently computed. An instance of these problems is congestion control in elastic traffic communication networks. For this problem, we proved that the  $\alpha$ -fair tradeoff policy achieves the optimal EJT for the special case of linear networks with uniform capacity. However, we also illustrated that the  $\alpha$ -fair tradeoff policy does not achieve the optimal EJT for general networks.

The conditions in Theorem 1 is also shown to be satisfied in the problem of radio resource allocation in the downlink of cellular networks. This enables us to devise computationally-efficient schedulers that achieve the optimal EJT in quasi-static and ergodic time-varying communication scenarios. Through analysis and extensive simulations, we compared the performance of the proposed optimal EJT scheduler,  $\alpha$ -fair scheduler, and MTMR scheduler. We found that the proposed optimal EJT scheduler outperforms both  $\alpha$ -fair and MTMR schedulers, in EJT, Efficiency-Median Tradeoff, and user satisfaction. On the other hand, the MTMR scheduler outperforms both the proposed optimal EJT and the  $\alpha$ -fair schedulers in Efficiency-Cell-edge Tradeoff; however, the MTMR scheduler incurs significant loss in EJT, Efficiency-Median Tradeoff, and user satisfaction.

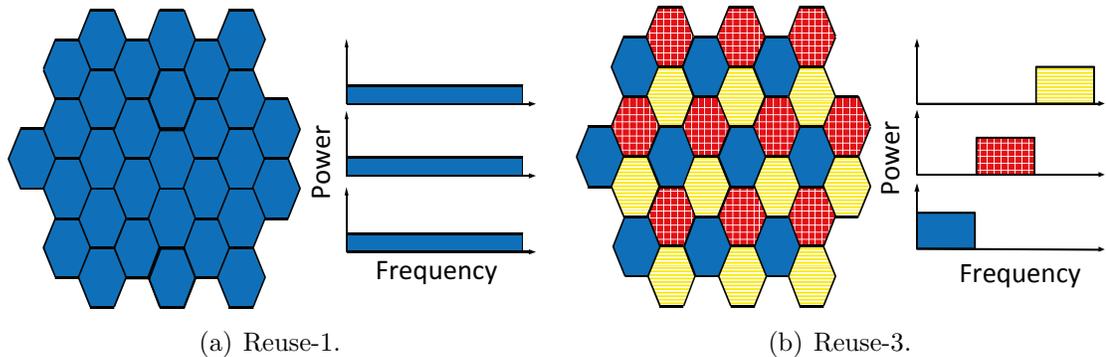
## Chapter 4

# Optimized Distributed Inter-cell Interference Coordination Scheme using Projected Subgradient and Network Flow Optimization

© 2011 IEEE. Parts of this chapter are reprinted, with permission, from:  
A. Bin Sediq, R. Schoenen, H. Yanıkömeroğlu, G. Senarath, and Z. Chao, “A novel distributed inter-cell interference coordination scheme based on projected subgradient and network flow optimizations,” in *Proc. IEEE International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC)*, pp. 1595–1600, September 2011.

### 4.1 Introduction

In order to achieve extremely high data rates in 4th generation (4G) and beyond-4G cellular networks, aggressive frequency reuse is inevitable due to the scarcity of the radio resources. Reuse-1, in which all radio resources are used in every sector, is an example of such an aggressive frequency reuse scheme (cf. Fig. 4.1(a)). While



**Figure 4.1:** Illustrations of reuse-1 and reuse-3.

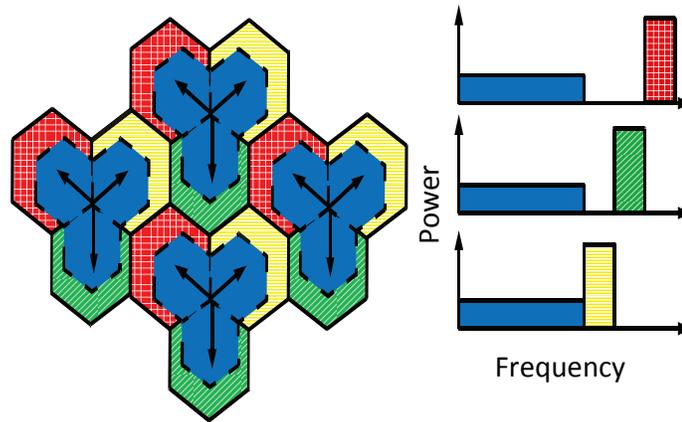
reuse-1 can potentially achieve high aggregate system throughput, it jeopardizes the throughput experienced by users close to the cell<sup>1</sup>-edge, due to the excessive interference experienced by these users. Therefore, it is vital for the network to use robust and efficient interference mitigation techniques.

Conventionally, interference is mitigated by static resource partitioning and frequency/sector planning, where close-by sectors are assigned orthogonal resources (clustering). A common example is reuse-3 (cluster size = 3 sectors) [40, Section 2.5], where adjacent sectors are assigned orthogonal channels, i.e., there is no inter-cell interference between adjacent sectors (cf. Fig. 4.1(b)). Although such techniques can reduce inter-cell interference and improve cell-edge user throughput, they suffer from two major drawbacks. First of all, the aggregate network throughput is significantly reduced since each sector has only a fraction of the available resources, which is equal to the reciprocal of the reuse factor. Secondly, conventional frequency/sector-planning may not be possible in emerging wireless networks where new multi-tier network elements (such as relays, femto-/pico-BSs, distributed antenna ports) are expected to be installed in an ad hoc manner in self-organizing networks (SON), without prior planning.

In order to reduce the effect of the first drawback, fractional frequency reuse

---

<sup>1</sup>The terms “cell” and “sector” are used interchangeably in this work.



**Figure 4.2:** Illustration of PFR, which is proposed in [42].

(FFR) schemes have been proposed. The key idea in FFR is to assign lower reuse factor for users near the cell-center and higher reuse factor for users at the cell-edge. The motivation behind such a scheme is that cell-edge users are more vulnerable to inter-cell interference than cell-center users. Soft Frequency Reuse (SFR) [41] and Partial Frequency Reuse (PFR) [42] (cf. Fig. 4.2) are two variations of FFR. A comparative study between reuse-1, reuse-3, PFR, and SFR is provided in [43, 44]. While FFR schemes recover some of the throughput loss due to partitioning, they require frequency/sector planning a priori, which is not desirable in future cellular networks as mentioned earlier. As a result, developing efficient dynamic ICIC schemes is vital to the success of future cellular networks.

One approach to tackle the ICIC problem is to devise adaptive FFR or SFR schemes. In [45], an adaptive FFR scheme is proposed where each BS chooses one of several reuse modes. A dynamic and centralized FFR scheme is proposed in [46] that outperforms conventional FFR schemes in terms of the total system throughput. In [47], the authors propose softer frequency reuse, which is a heuristic algorithm based on modifying the proportional fair algorithm and the SFR scheme. In [48], a heuristic algorithm based on adaptive SFR is proposed for the uplink. In [49], the authors propose gradient-based distributed schemes that create SFR patterns, in an

effort to achieve local maximization of the network utility.

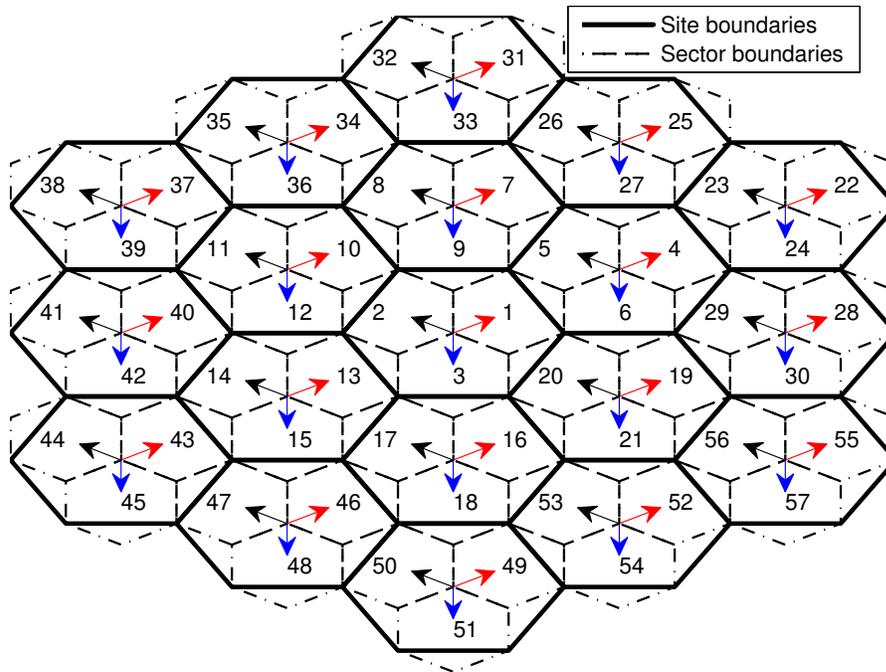
Another approach to tackle the ICIC problem is to remove the limitations imposed by the FFR schemes and view the ICIC problem as a multi-cell scheduling problem. In [50], the authors propose suboptimal semi-distributed radio resource control scheme where a radio network controller is assumed to be connected to all BSs. The authors conclude that the problem is NP-hard and they resort to heuristics that show improvement in the performance. In [51], a graph-theoretic approach is taken to develop an ICIC scheme in which information about the interference experienced by each user terminal (UT) is inferred from the diversity set of that UT. In [52], a game-theoretic approach is pursued and a decentralized algorithm is developed. The proposed algorithm converges to a Nash equilibrium in a simplified cellular system. Nevertheless, a significant gap is observed between the proposed algorithm and the globally optimum one, which is computationally demanding. In [53], a partly distributed two-level ICIC scheme is proposed where a centralized entity is required to solve a binary linear optimization problem, which is generally not solvable in polynomial time.

The focus of this chapter is to overcome the drawbacks of the existing schemes mentioned above and the main contributions can be summarized as follows:

1. A novel ICIC algorithm is proposed. The proposed algorithm can be distinguished from the existing algorithms in literature as follows. Unlike the partly centralized algorithms proposed in [50, 53–57], our algorithm is fully distributed and it does not require a central controller. Moreover, the proposed algorithm can be used with any AMC scheme, including discrete-rate AMC schemes, unlike the schemes in [49, 54] which require the AMC function to be differentiable. While there are distributed algorithms and can be used with any AMC, such as those in [58–61], such algorithms are developed based on heuristics and intuition. Hence, it is unclear how close these heuristics to the optimum algorithm. Also,

heuristics provide little insight on the problem dynamics. We acknowledge the difficulty of solving the ICIC problem optimally; indeed, the optimization problem that we consider herein is strongly NP-hard. However, rather than pursuing a heuristic approach, we carefully approximate the problem, and through rigorous mathematical derivations, we devise a polynomial-time distributed scheme. Although the scheme is developed solely based on mathematical derivations, it has simple and intuitive interpretation which would shed light on how distributed ICIC schemes can be designed. Simulation results show that the proposed scheme achieves a performance that is close to the optimal one achieved by exhaustive search, which validates our approximation. Furthermore, extensive simulation results show that the proposed scheme achieves significantly higher cell-edge throughput, higher aggregate throughput, and lower outage probability as compared to the baseline static and dynamic schemes. Although the proposed scheme is presented herein in the context of a homogeneous network, where the access network consists of macro-BSs, it can also work as an enhanced ICIC (eICIC) scheme [62] in the context of a heterogeneous network, where the access network consists of macro-BSs, pico-BSs, femto-BSs, and relays, as we will show in Chapter 5.

2. We realize that the devised optimization problem has a minimum cost network flow structure. This is an important insight to the ICIC optimization problem because by exploiting such structure, we can use network-based algorithms which have significantly reduced complexity as compared to the general-purpose convex or linear optimization algorithms [4, p. 402].



**Figure 4.3:** The considered network layout which consists of 19 cell sites and 3 sectors per site.

## 4.2 System Model

We consider the network model described by the IMT-Advanced evaluation guidelines [34]. Based on these guidelines, the considered network consists of  $K$  sectors served by  $K/3$  BSs as shown in Fig. 4.3. Each BS is equipped with a tri-sector antenna to serve a cell-site that consists of 3 sectors. Each BS can communicate with its neighboring BSs; this is supported in most cellular network standards, e.g., using the R8 interface in IEEE 802.16m standard and the X2 interface in LTE and LTE-advanced standards. Hexagonal sectors are considered herein according to IMT-Advanced guidelines; nevertheless, the proposed scheme works also for arbitrary sector shapes. We focus on the downlink scenario in this thesis.

Orthogonal Frequency Division Multiple Access (OFDMA) is used as the multiple access scheme, since it is adopted in most of the contemporary cellular standards.

Among the advantages offered by OFDMA is its scheduling flexibility, since users can be scheduled in both time and frequency, which can be exploited to gain time, frequency, and multi-user diversity. The time and frequency radio resources are grouped into time-frequency resource blocks (RBs). RB is the smallest radio resource block that can be scheduled to a UT and it consists of  $N_s$  OFDM symbols in the time dimension and  $N_f$  sub-carriers in the frequency dimension. The total number of RBs is denoted by  $N$ . The number of UTs in sector  $k$  is denoted by  $M^{(k)}$ . Both the BSs and the UTs are assumed to have single antenna each. Similar to [53], we assume that each UT estimates and reports to its serving BS the channel from its serving sector's antenna and from the first-tier interfering sectors. The SINR observed by UT  $m \in \mathcal{M}^{(k)} \triangleq \{1, \dots, M^{(k)}\}$  in sector  $k \in \mathcal{K} \triangleq \{1, \dots, K\}$  on RB  $n \in \mathcal{N} \triangleq \{1, \dots, N\}$  is given by [53]

$$\Gamma_{m,n}^{(k)} = \frac{P_C H_{m,n}^{(k,k)}}{P_C \sum_{\tilde{k}=1, \tilde{k} \neq k}^K (1 - I_n^{(\tilde{k})}) H_{m,n}^{(k,\tilde{k})} + P_N}, \quad (4.1)$$

where  $P_C$  represents the transmitted power per RB,  $P_N$  represents the thermal noise power per RB,  $I_n^{(k)}$  is a binary variable indicating whether RB  $n$  is restricted in sector  $k$  ( $I_n^{(k)} = 1$ ) or not ( $I_n^{(k)} = 0$ ), and  $H_{m,n}^{(k,\tilde{k})}$  represents the channel gain from sector  $\tilde{k}$  on RB  $n$  to UT  $m$  served by sector  $k$ . Large scale channel variations (due to distance-dependant attenuation and shadowing), antenna gains, and multipath fading are all captured in  $H_{m,n}^{(k,\tilde{k})}$ . The achievable rate on RB  $n$  of UT  $m$  in sector  $k$  is given by

$$R_{m,n}^{(k)} = f(\Gamma_{m,n}^{(k)}) \quad (\text{bit/sec}), \quad (4.2)$$

where  $f(\cdot)$  is the AMC function that maps SINR to rate. The function  $f(\cdot)$  is assumed to be nondecreasing, possibly discontinuous, which is the case for all practical AMC schemes.

## 4.3 Problem Statement

The ICIC problem can be considered as network-wide joint scheduling and power control problem. In the following, we start by reviewing the literature on power control in Section 4.3.1. Based on the literature survey, we find that on-off power control, where each RB is either transmitted with a total power of  $P_C$  or not transmitted at all, is an attractive option from both tractability point of view and from practical implementation. As a result, we formulate the ICIC problem as a network-wide joint scheduling and binary power control in Section 4.3.2.

### 4.3.1 Power Control

For single-cell case, power allocation<sup>2</sup> is well-understood. For a single user, the well-known water-filing algorithm is an efficient algorithm to maximize the total throughput by allocating the total power over the RBs [63]. For single-cell multiuser case, power allocation and scheduling can be optimized jointly [12, 64–66]. In [67], it is proven that, under total power constraints, the sum throughput is maximized when each RB is assigned to only the user which has the best channel on that RB. Optimum power can be solved using conventional water-filling per user. It is shown that optimum power distribution provides only marginal gain as compared to equal power scheme. As a result, equal-power allocation is widely used in single-cell case.

For the multi-cell case, power control is more complex than the case of a single-cell since interference has to be taken into account. The resulted optimization problem in this case is not convex, and in general, hard to solve. In [68, 69], maximizing the weighted-sum rate for single-carrier system and single-user per cell was formulated as a geometric programming problem based on high SINR approximations. Such formulation can be readily transformed into an equivalent convex optimization problem and

---

<sup>2</sup>It is common in the literature to use the term “power allocation” for the single-cell case and the term “power control” for the multi-cell case.

as a result, the globally optimum solution based on the high SINR approximation can be efficiently computed in distributed manner. In the low to medium SINR regime, a heuristic based on successive convex optimization techniques was developed, without guarantee on the global optimality. The authors in [70] proposed a very simple distributed scheme that solve the same high-SINR problem in [68,69], much faster than using geometric programming. The key idea is to use the fixed-point theorem and the technique of standard interference function that was provided in the seminal work of Yates in [71] (see also [72] for a through treatment of the standard interference function). The solution given in [68,69] is only accurate in the high-SINR regime [73]. Indeed, as observed in [74], the high-SINR approximation may lead to solutions that are far from optimum. High SINR approximation also doesn't allow the BS to turn off the power completely, which may again lead to solutions away from the optimum solution [75]. In [74], maximizing total throughput is formulated as linear multiplicative fractional programming. Although global optimal power allocation can be achieved in a general SINR regime, polynomial-time convergence is not guaranteed.

In [76,77], solving the power control in multi-cell multi-carrier system and single user per cell, under sum power constraint, is tackled by maximizing the weighted-sum rate in the dual domain. In [76], a sufficient condition, called time-sharing condition, is presented and when it is satisfied, then the duality gap is zero, i.e., both the dual and primal optimization problem have the same optimal value. It is also shown that this time-sharing property is satisfied as the number of RBs go to infinity. In [78], it is shown that this condition is also satisfied for multiple users per cell. While the dual optimization problem is convex and can be solved efficiently using subgradient method, constructing the dual problem itself requires exhaustive search which has exponential complexity. Moreover, only centralized algorithms are presented in [78] and they are not necessarily globally optimum as they depend on the initial starting point.

In [79], the computational complexity of maximizing different utilities is studied, for different number of RBs and different number of cells. An important result in this work is that maximizing the weighted sum-rate is strongly NP-hard even for single RB and single user per cell, meaning that it not only hard to solve, but also hard to obtain approximate optimal solution with guaranteed optimality gap. This explains the reasons why all the existing algorithms that find the global optimum solution that maximize weighted sum-rate, even for the simplest case of single RB and single user per cell, have worst-case exponential complexity, see e.g. [80–83].

To simplify the distributed implementation of power control, binary power control (including on-off power control), i.e., either transmitting with the minimum power or the maximum power, is considered in [84–88]. Intuitively, binary power control is a special case of continuous power control and thus it is expected to be suboptimal. Interestingly, binary power control is shown to be optimal or near-optimal in special scenarios. For instance, in [84, 85], it is shown that the optimal power allocation for the case of two sectors is binary power control (including on-off power control). For more than two sectors, it is shown in [73] that binary power control is no longer optimum in general, although simulation results for a small number of sectors hints that it is near-optimum. Moreover, binary power control is optimum in the low-SINR regime or when the all SINRs are equal, for any number of sectors. Moreover, it was shown in [86] that binary power control is optimum for a symmetric interference channel, where all direct channel gains are assumed to be equal to some value and all cross channel gains are assumed to be equal to another value. In [87, 88], binary power control is also shown to be optimal for the uplink power control for a single cell, multi-users, and single RB. Motivated by the simplicity of implementing on-off power control, we consider on-off power control hereinafter. In particular, we assume each RB can be either used with full power or not used at all in each cell.

### 4.3.2 Problem Formulation

Let us consider a generic scheduler implemented in sector  $k$ , without ICIC, such that it maximizes the weighted sum of the UTs' rates in sector  $k$ . This scheduler is implemented by solving the following optimization problem for each RB  $n$  in every sub-frame  $t$ :

$$\underset{\{\rho_{m,n}^{(k)}\}}{\text{maximize}} \quad \sum_{m=1}^{M^{(k)}} w_m^{(k)}(t) \rho_{m,n}^{(k)}(t) R_{m,n}^{(k)}(t) \quad (4.3a)$$

$$\text{subject to} \quad \sum_{m=1}^{M^{(k)}} \rho_{m,n}^{(k)}(t) = 1, \quad (4.3b)$$

$$\rho_{m,n}^{(k)}(t) \in \{0, 1\}, \quad \forall m \in \mathcal{M}^{(k)}, \quad (4.3c)$$

where  $w_m^{(k)}(t)$  is the weight of UT  $m$  in sector  $k$ ,  $\{\rho_{m,n}^{(k)}(t)\}$  are the binary decision variables such that  $\rho_{m,n}^{(k)}(t) = 1$  when RB  $n$  is assigned to UT  $m$  in sector  $k$ , and  $\rho_{m,n}^{(k)}(t) = 0$  otherwise. Constraint (4.3b) ensures that each RB is assigned to only one user in sector  $k$ .

The scheduler described above can be used to control the desired fairness-throughput tradeoff in the long-term average rates by updating the weights  $\{w_m^{(k)}(t)\}$  in every sub-frame using the gradient scheduling algorithm explained in Section 3.5. For instance, for the  $\alpha$ -fair and optimal EJT scheduling,  $\{w_m^{(k)}(t)\}$  can be calculated as

$$w_m^{(k)}(t) = \begin{cases} (W_m^{(k)}(t-1))^{-\alpha}, & \alpha\text{-fairness,} \\ \beta - W_m^{(k)}(t-1), & \text{Optimal EJT,} \end{cases} \quad (4.4)$$

where  $\alpha$  and  $\beta$  are tradeoff parameters as explained in Chapters 2 and 3, and  $W_m^{(k)}(t)$  denotes the average rate of user  $m$  in sector  $k$  in sub-frame  $t$ , averaged using

exponentially-weighted moving average with a small smoothing factor  $\mu \in (0, 1)$ , i.e.,

$$\begin{aligned} W_m^{(k)}(t) &= \mu \sum_{i=0}^t (1 - \mu)^{i-t} \sum_{n=1}^N \rho_{m,n}^{(k)}(i) R_{m,n}^{(k)}(i) \\ &= (1 - \mu) W_m^{(k)}(t - 1) + \mu \sum_{n=1}^N \rho_{m,n}^{(k)}(t) R_{m,n}^{(k)}(t). \end{aligned} \quad (4.5)$$

Another approach to vary the degree of fairness-throughput tradeoff is by imposing a minimum long-term average rate ( $\bar{R}_{\min}$ ), varying  $\bar{R}_{\min}$  [6], and updating  $\{w_m^{(k)}(t)\}$  according to the algorithm proposed in [33] (see also Section 3.5.1). We investigate all these approaches in this Chapter.

It can be shown that the solution to (4.3) is given by

$$\rho_{m,n}^{*(k)}(t) = \begin{cases} 1, & m = \arg \max_{m \in \mathcal{M}^{(k)}} w_m^{(k)}(t) R_{m,n}^{(k)}(t), \\ & \text{and } \max_{m \in \mathcal{M}^{(k)}} w_m^{(k)}(t) R_{m,n}^{(k)}(t) \geq 0, \\ 0, & \text{otherwise.} \end{cases} \quad (4.6)$$

To simplify the notation, we drop the sub-frame index ( $t$ ) hereinafter.

In order to improve the performance of the above mentioned scheduler, we seek an ICIC scheme that coordinates the scheduling in all sectors in order to maximize the weighted sum of the rates of all users in the network. The optimization problem for each RB  $n$  to attain this goal can be formulated as

$$\begin{aligned} & \underset{\{\rho_{m,n}^{(k)}\}, \{I_n^{(k)}\}}{\text{maximize}} && \sum_{k=1}^K \sum_{m=1}^{M^{(k)}} w_m^{(k)} \rho_{m,n}^{(k)} R_{m,n}^{(k)} \end{aligned} \quad (4.7a)$$

$$\text{subject to} \quad \sum_{m=1}^{M^{(k)}} \rho_{m,n}^{(k)} = 1 - I_n^{(k)}, \quad \forall k \in \mathcal{K}, \quad (4.7b)$$

$$\rho_{m,n}^{(k)}, I_n^{(k)} \in \{0, 1\}, \quad \forall k \in \mathcal{K}, m \in \mathcal{M}^{(k)}, \quad (4.7c)$$

where the binary variables  $\{I_n^{(k)}\}$  are introduced such that  $I_n^{(k)} = 1$  when the use of RB  $n$  is restricted in sector  $k$  and  $I_n^{(k)} = 0$  otherwise. Constraint (4.7b) ensures that each RB  $n$  is assigned to only one UT in sector  $k$ , given that RB  $n$  is not restricted in sector  $k$ .

The network-wide optimization problem in (4.7) is difficult to solve for the following reasons. First of all, this problem belongs to the class of non-linear binary combinatorial optimization problems ( $R_{m,n}^{(k)}$  is a nonlinear function of  $I_n^{(k)}$ , cf. (4.1) and (4.2)), which are generally difficult to solve in polynomial time. Indeed, (4.7) is strongly NP-hard even for the case of a single user per sector [79]. This means that not only (4.7) is hard to solve in polynomial time, but it is also hard to find an approximation algorithm with guaranteed optimality gap. Moreover, the objective function is dependant on the AMC strategy that is used. Hence, an optimal solution for a given AMC strategy may not be optimal for another AMC strategy. Since AMC strategies are operator dependant, it is desirable to develop an algorithm that is independent of the chosen AMC strategy. Finally, it is desirable to solve (4.7) in a distributed manner since most contemporary standards (such as LTE, LTE-Advanced, IEEE 802.16m) do not support a central controller. The main contribution of this chapter is to circumvent these difficulties and develop an efficient algorithm that obtain a near-optimal solution of (4.7). The development of such an algorithm is explained in the following section.

## 4.4 Proposed Algorithm

In this section, we show the steps used to develop the proposed algorithm. We start in Section 4.4.1 by introducing a bound and argue that this bound is a good metric for optimization. Using the bound, we get a binary non-linear optimization problem and we transform it into an equivalent binary linear optimization problem in

Section 4.4.2. Next, we relax the resulted binary linear optimization problem into a linear optimization problem in Section 4.4.3. Then, we devise a distributed algorithm in Section 4.4.4 by decomposing the problem into a master problem and multiple subproblems using the primal-decomposition method. To reduce the computational complexity, we show in Section 4.4.5 that each subproblem is amenable to powerful network flow optimization methods. We then present a pseudocode of the proposed algorithm in Section 4.4.6 and discuss its complexity in Section 4.4.8. A block diagram that shows the roadmap of this section is shown in Fig. 4.4.

### 4.4.1 Bound Optimization

The SINR expression in (4.1) can be lower-bounded as

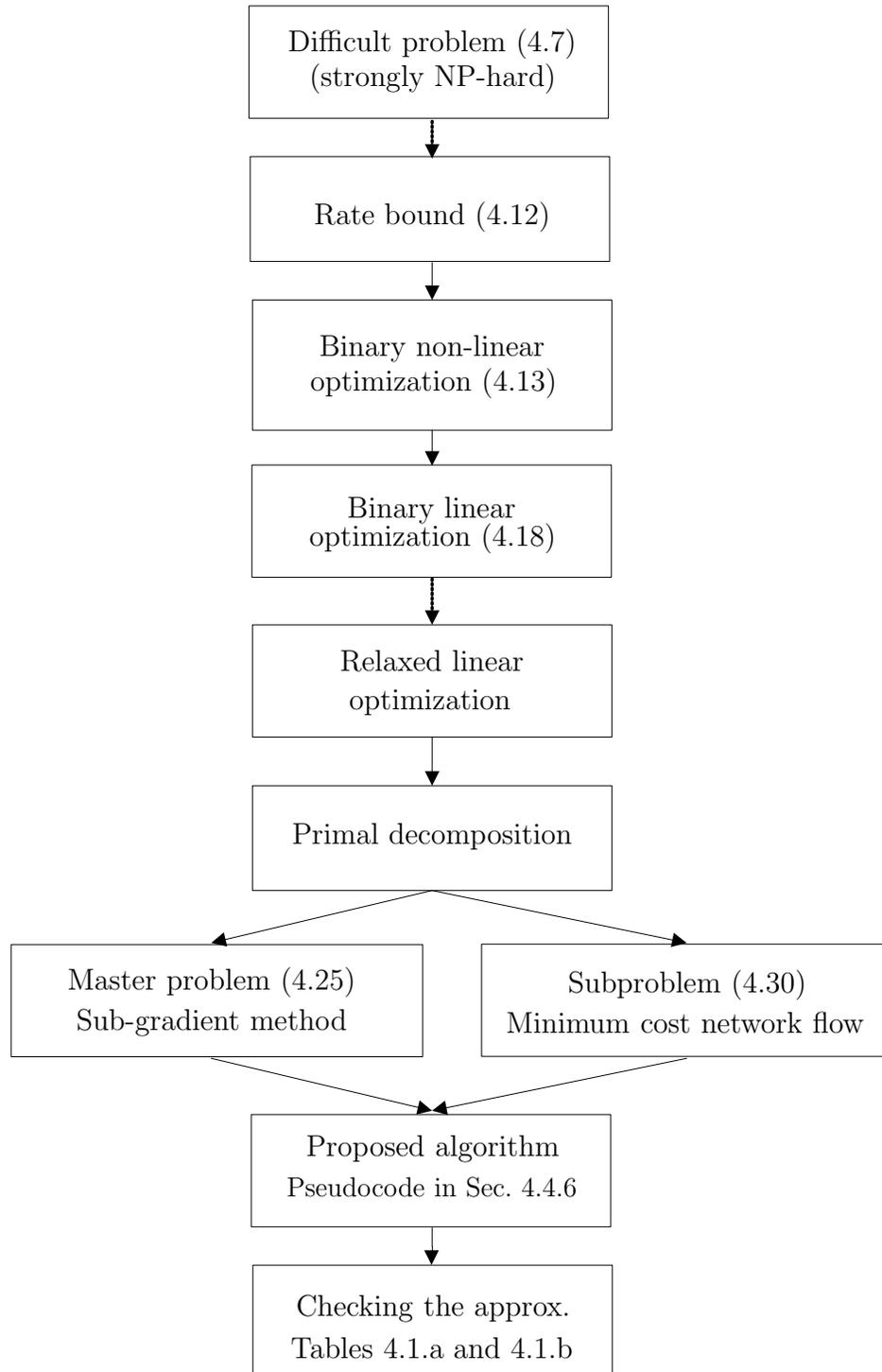
$$\Gamma_{m,n}^{(k)} \geq \frac{P_C H_{m,n}^{(k,k)}}{P_C \sum_{\tilde{k}=1, \tilde{k} \neq k}^K H_{m,n}^{(k,\tilde{k})} - \max_{\tilde{k} \in \mathcal{K}^{(k)}} I_n^{(\tilde{k})} P_C H_{m,n}^{(k,\tilde{k})} + P_N}, \quad (4.8)$$

where  $\mathcal{K}^{(k)}$  is the set of indices of the neighboring sectors of sector  $k$ , which are considered in this work to be the 6 first-tier interfering sectors seen by sector  $k$ ; e.g., in Fig. 4.3,  $\mathcal{K}^{(3)} = \{1, 2, 13, 17, 16, 20\}$ <sup>3</sup>. The bound in (4.8) is obtained by assuming that RB  $n$  is used in all sectors, except at most one sector. When there are more than one sector restricted from using RB  $n$ , only the most dominant restricted interferer is considered<sup>4</sup>. This bound is exact if the number of restricted interferers is less or equal to one and it is tight for small number of restricted interferers. This is a good bound to be used for optimization (maximization) for the following reasons. First of

---

<sup>3</sup>We assume a wraparound layout so each sector has 6 first-tier interfering sectors; e.g.,  $\mathcal{K}^{(25)} = \{26, 27, 23, 51, 50, 48\}$ .

<sup>4</sup>Note that this bound considers the most dominant *restricted* interferer which is not necessarily the most dominant interferer when the most dominant interferer is not restricted (i.e., when it is not turned off). This differentiates (4.8) from the expression considered in [50], where the most dominant interferer is considered.



**Figure 4.4:** A block diagram that shows the steps used to develop the proposed algorithm.

all, if the bound is increased by  $\Delta$ , then the exact expression given by (4.1) will also increase by at least  $\Delta$ , as provided in the following Lemma.

**Lemma 4.** *The SINR bound given by (4.8) is tight in the following sense:*

*if the bound is increased by  $\Delta$ , then the actual SINR will increase by at least  $\Delta$ .*

*Proof.* Let  $\tilde{k}^* = \arg \max_{\tilde{k} \in \mathcal{K}^{(k)}} I_n^{(\tilde{k})} H_{m,n}^{(k,\tilde{k})}$ , then the exact SINR expression given by (4.1) can be written as

$$\begin{aligned} \Gamma_{m,n}^{(k)} &= \frac{P_C H_{m,n}^{(k,k)}}{P_C \left( \sum_{\tilde{k}=1, \tilde{k} \neq k}^K H_{m,n}^{(k,\tilde{k})} - I_n^{\tilde{k}^*} H_{m,n}^{(k,\tilde{k}^*)} - \sum_{\tilde{k}=1, \tilde{k} \neq k, \tilde{k} \neq \tilde{k}^*}^K I_n^{(\tilde{k})} H_{m,n}^{(k,\tilde{k})} \right) + P_N}, \\ &= \frac{P_C H_{m,n}^{(k,k)}}{P_C \left( \sum_{\tilde{k}=1, \tilde{k} \neq k}^K H_{m,n}^{(k,\tilde{k})} - I_n^{\tilde{k}^*} H_{m,n}^{(k,\tilde{k}^*)} \right) + P_N} \times \left( 1 + \frac{P_C \sum_{\tilde{k}=1, \tilde{k} \neq k, \tilde{k} \neq \tilde{k}^*}^K I_n^{(\tilde{k})} H_{m,n}^{(k,\tilde{k})}}{P_C \sum_{\tilde{k}=1, \tilde{k} \neq k}^K \left( 1 - I_n^{(\tilde{k})} \right) H_{m,n}^{(k,\tilde{k})} + P_N} \right), \quad (4.9) \\ &\geq \frac{P_C H_{m,n}^{(k,k)}}{P_C \left( \sum_{\tilde{k}=1, \tilde{k} \neq k}^K H_{m,n}^{(k,\tilde{k})} - I_n^{\tilde{k}^*} H_{m,n}^{(k,\tilde{k}^*)} \right) + P_N}, \end{aligned}$$

where the right-hand side of the inequality is the bound given by (4.8). From the previous expression, if the bound is increased by  $\Delta$ , then  $\Gamma_{m,n}^{(k)}$  will increase by  $\Delta \times \left( 1 + \frac{P_C \sum_{\tilde{k}=1, \tilde{k} \neq k, \tilde{k} \neq \tilde{k}^*}^K I_n^{(\tilde{k})} H_{m,n}^{(k,\tilde{k})}}{P_C \sum_{\tilde{k}=1, \tilde{k} \neq k}^K \left( 1 - I_n^{(\tilde{k})} \right) H_{m,n}^{(k,\tilde{k})} + P_N} \right)$ , i.e.,  $\Gamma_{m,n}^{(k)}$  will increase by at least  $\Delta$ .  $\square$

In addition to the tightness of the bound, it is already observed in the literature that restricting more than two sectors from using a particular RB can degrade the overall system performance (cf. [53]). Finally, and most importantly, based on this bound, we can develop a distributed optimization framework that is applicable to a wide range of schedulers and AMC strategies. This framework can be implemented very efficiently, and can achieve near-optimal performance, as we will see later.

Using the SINR bound given in (4.8), we now construct a similar bound on the rates,  $\{R_{m,n}^{(k)}\}$ , given in (4.2). To do so, we define  $\gamma_{m,n}^{(k)}$  and  $r_{m,n}^{(k)}$  as the SINR and the corresponding achievable rate on RB  $n$  of UT  $m$  in sector  $k$ , if all sectors use RB  $n$ ,

i.e.,

$$\gamma_{m,n}^{(k)} = \frac{P_C H_{m,n}^{k,k}}{P_C \sum_{\tilde{k} \neq k} H_{m,n}^{(k,\tilde{k})} + P_N}, r_{m,n}^{(k)} = f\left(\gamma_{m,n}^{(k)}\right). \quad (4.10)$$

Similarly, we define  $\tilde{\gamma}_{m,n}^{(k,\tilde{k})}$  and  $\tilde{r}_{m,n}^{(k,\tilde{k})}$  as the SINR and the achievable rate on RB  $n$  of UT  $m$  in sector  $k$ , if only sector  $\tilde{k} \in \mathcal{K}^{(k)}$  is restricted from using RB  $n$ , i.e.,

$$\tilde{\gamma}_{m,n}^{(k,\tilde{k})} = \frac{P_C H_{m,n}^{(k,k)}}{P_C \sum_{\hat{k} \neq k} H_{m,n}^{k,\hat{k}} - P_C H_{m,n}^{(k,\tilde{k})} + P_N}, \tilde{r}_{m,n}^{(k,\tilde{k})} = f\left(\tilde{\gamma}_{m,n}^{(k,\tilde{k})}\right) - r_{m,n}^{(k)}. \quad (4.11)$$

Using (4.10) and (4.11), a bound on the rates,  $\{R_{m,n}^{(k)}\}$ , given by (4.2) is constructed based on the SINR bound given by (4.8) as

$$R_{m,n}^{(k)} \geq r_{m,n}^{(k)} + \max_{\tilde{k} \in \mathcal{K}^{(k)}} I_n^{(\tilde{k})} \tilde{r}_{m,n}^{(k,\tilde{k})}. \quad (4.12)$$

By substituting (4.12) in (4.7), the optimization problem can be approximated as

$$\text{maximize}_{\{\rho_{m,n}^{(k)}\}, \{I_n^{(k)}\}} \sum_{k=1}^K \sum_{m=1}^{M^{(k)}} w_m^{(k)} \rho_{m,n}^{(k)} \left( r_{m,n}^{(k)} + \max_{\tilde{k} \in \mathcal{K}^{(k)}} I_n^{(\tilde{k})} \tilde{r}_{m,n}^{(k,\tilde{k})} \right) \quad (4.13a)$$

$$\text{subject to} \quad \sum_{m=1}^{M^{(k)}} \rho_{m,n}^{(k)} = 1 - I_n^{(k)}, \quad \forall k \in \mathcal{K}, \quad (4.13b)$$

$$\rho_{m,n}^{(k)}, I_n^{(k)} \in \{0, 1\}, \quad \forall k \in \mathcal{K}, m \in \mathcal{M}^{(k)}. \quad (4.13c)$$

The optimization problem (4.13) is a non-linear binary integer optimization problem, which is in general, difficult to solve. As an intermediate step to reduce the complexity of solving (4.13), we convert it to an equivalent binary linear optimization problem in the following section.

## 4.4.2 Transforming (4.13) Into an Equivalent Binary Linear Optimization Problem

Our approach to tackle the binary non-linear optimization problem given by (4.13) is to first transform it into an equivalent binary linear optimization problem. While binary linear optimization problems are still not easy to solve in general, good approximate solutions can be computed efficiently by relaxing the constraints that the variables have to be binary which allows using real variables instead. The challenge is to construct an equivalent binary linear optimization problem that has a tight relaxation which means that the solution obtained by solving the relaxed version is very close to the one obtained by solving the binary linear optimization problem. This challenge is addressed in this section.

In order to convert the binary non-linear optimization problem given by (4.13) into an equivalent binary linear optimization problem, we need to convert the term  $\rho_{m,n}^{(k)} \max_{\tilde{k} \in \mathcal{K}^{(k)}} I_n^{(\tilde{k})} \tilde{r}_{m,n}^{(k,\tilde{k})}$  into a linear term. There are two sources of non-linearity in this term: the point-wise maximum and the multiplication. The non-linear term can be written as

$$\rho_{m,n}^{(k)} \max_{\tilde{k} \in \mathcal{K}^{(k)}} I_n^{(\tilde{k})} \tilde{r}_{m,n}^{(k,\tilde{k})} = \max_{\tilde{\rho}_{m,n}^{(k,\tilde{k})} \in \mathcal{S}} \sum_{\tilde{k} \in \mathcal{K}^{(k)}} \tilde{\rho}_{m,n}^{(k,\tilde{k})} \tilde{r}_{m,n}^{(k,\tilde{k})}, \quad (4.14)$$

where the variables  $\{\tilde{\rho}_{m,n}^{(k,\tilde{k})}\}$  are introduced as auxiliary variables that facilitate the conversion of the non-linear term into a linear term and

$$\begin{aligned}
\mathcal{S} = \left\{ \tilde{\rho}_{m,n}^{(k,\tilde{k})} : \sum_{\tilde{k} \in \mathcal{K}^{(k)}} \tilde{\rho}_{m,n}^{(k,\tilde{k})} \leq 1, \quad \forall k \in \mathcal{K}, m \in \mathcal{M}^{(k)} \right. \\
\tilde{\rho}_{m,n}^{(k,\tilde{k})} \leq \rho_{m,n}^{(k)}, \quad \forall k \in \mathcal{K}, \tilde{k} \in \mathcal{K}^{(k)}, m \in \mathcal{M}^{(k)} \\
\tilde{\rho}_{m,n}^{(k,\tilde{k})} \leq I_n^{(\tilde{k})}, \quad \forall k \in \mathcal{K}, \tilde{k} \in \mathcal{K}^{(k)}, m \in \mathcal{M}^{(k)} \\
\left. \tilde{\rho}_{m,n}^{(k,\tilde{k})} \in \{0, 1\}, \quad \forall k \in \mathcal{K}, \tilde{k} \in \mathcal{K}^{(k)}, m \in \mathcal{M}^{(k)} \right\}.
\end{aligned} \tag{4.15}$$

To see the equivalence, we note that since all variables assume binary values, then the point-wise maximum in the originally non-linear term is captured in the objective function and the first inequality in (4.15). Moreover, the multiplication in the originally non-linear term is captured in the second and the third inequalities in (4.15).

By replacing the term  $\rho_{m,n}^{(k)} \max_{\tilde{k} \in \mathcal{K}^{(k)}} I_n^{(\tilde{k})} \tilde{r}_{m,n}^{(k,\tilde{k})}$  in (4.13a) with (4.14) and (4.15), we get an equivalent binary linear optimization problem. Unfortunately, we found experimentally that such an equivalent optimization problem leads to loose linear relaxation, i.e., it results in solutions that are far from the binary optimal solutions. As a result, we seek tighter equivalent formulations.

A general approach to get tighter relaxation is by adding additional constraints that are called valid constraints [89, p. 585]. These valid constraints do not change the set of feasible binary solutions; however, with proper choice of these valid constraints, one can obtain tighter relaxations. Based on this approach, we construct a set  $\mathcal{S}'$  that is equivalent to  $\mathcal{S}$  for the considered binary optimization problem, by adding two valid constraints:

$$\begin{aligned}
\sum_{\tilde{k} \in \mathcal{K}^{(k)}} \tilde{\rho}_{m,n}^{(k,\tilde{k})} \leq \rho_{m,n}^{(k)}, \quad \forall k \in \mathcal{K}, m \in \mathcal{M}^{(k)} \\
\sum_{m=1}^{M^{(\tilde{k})}} \tilde{\rho}_{m,n}^{(k,\tilde{k})} \leq I_n^{(\tilde{k})}, \quad \forall k \in \mathcal{K}, \tilde{k} \in \mathcal{K}^{(k)}.
\end{aligned} \tag{4.16}$$

Hence,  $\mathcal{S}'$  is given by

$$\begin{aligned} \mathcal{S}' = \left\{ \tilde{\rho}_{m,n}^{(k,\tilde{k})} : \sum_{\tilde{k} \in \mathcal{K}^{(k)}} \tilde{\rho}_{m,n}^{(k,\tilde{k})} \leq \rho_{m,n}^{(k)}, \quad \forall k \in \mathcal{K}, m \in \mathcal{M}^{(k)} \right. \\ \sum_{m=1}^{M(\tilde{k})} \tilde{\rho}_{m,n}^{(k,\tilde{k})} \leq I_n^{(\tilde{k})}, \quad \forall k \in \mathcal{K}, \tilde{k} \in \mathcal{K}^{(k)} \\ \left. \tilde{\rho}_{m,n}^{(k,\tilde{k})} \in \{0, 1\}, \quad \forall k \in \mathcal{K}, \tilde{k} \in \mathcal{K}^{(k)}, m \in \mathcal{M}^{(k)} \right\}. \end{aligned} \quad (4.17)$$

Note that the constraints  $\tilde{\rho}_{m,n}^{(k,\tilde{k})} \leq \rho_{m,n}^{(k)}$  and  $\tilde{\rho}_{m,n}^{(k,\tilde{k})} \leq I_n^{(\tilde{k})}$ , given originally in (4.15), are omitted from the definition of  $\mathcal{S}'$  given in (4.17) since these constraints define a subset of the first two constraints given in (4.17). The equivalence of  $\mathcal{S}$  and  $\mathcal{S}'$  is provided in the following Lemma.

**Lemma 5.** *The sets  $\mathcal{S}$  and  $\mathcal{S}'$  given by (4.15) and (4.17), respectively, are equivalent.*

*Proof.* We proceed by proving that  $\mathcal{S} \subseteq \mathcal{S}'$  and  $\mathcal{S}' \subseteq \mathcal{S}$ . To prove that  $\mathcal{S} \subseteq \mathcal{S}'$ , we assume that  $\tilde{\rho}_{m,n}^{(k,\tilde{k})} \in \mathcal{S}$  and deduce that  $\tilde{\rho}_{m,n}^{(k,\tilde{k})} \in \mathcal{S}'$ . Since  $\sum_{\tilde{k} \in \mathcal{K}^{(k)}} \tilde{\rho}_{m,n}^{(k,\tilde{k})} \leq 1$  and  $\tilde{\rho}_{m,n}^{(k,\tilde{k})} \leq \rho_{m,n}^{(k)}$ , then  $\sum_{\tilde{k} \in \mathcal{K}^{(k)}} \tilde{\rho}_{m,n}^{(k,\tilde{k})} \leq \rho_{m,n}^{(k)}$ . Moreover, since  $\tilde{\rho}_{m,n}^{(k,\tilde{k})} \leq \rho_{m,n}^{(k)}$ , then  $\sum_{m=1}^{M(k)} \tilde{\rho}_{m,n}^{(k,\tilde{k})} \leq \sum_{m=1}^{M(k)} \rho_{m,n}^{(k)} \leq 1$ . Since  $\sum_{m=1}^{M(k)} \tilde{\rho}_{m,n}^{(k,\tilde{k})} \leq 1$ ,  $\tilde{\rho}_{m,n}^{(k,\tilde{k})} \leq I_n^{(\tilde{k})}$ , and  $\tilde{\rho}_{m,n}^{(k,\tilde{k})} \in \{0, 1\}$ , then  $\sum_{m=1}^{M(k)} \tilde{\rho}_{m,n}^{(k,\tilde{k})} \leq I_n^{(\tilde{k})}$ . Consequently,  $\tilde{\rho}_{m,n}^{(k,\tilde{k})} \in \mathcal{S}'$ , which means that  $\mathcal{S} \subseteq \mathcal{S}'$ .

Similarly, to show that  $\mathcal{S}' \subseteq \mathcal{S}$ , we assume that  $\tilde{\rho}_{m,n}^{(k,\tilde{k})} \in \mathcal{S}'$  and deduce that  $\tilde{\rho}_{m,n}^{(k,\tilde{k})} \in \mathcal{S}$ . Since  $\sum_{\tilde{k} \in \mathcal{K}^{(k)}} \tilde{\rho}_{m,n}^{(k,\tilde{k})} \leq \rho_{m,n}^{(k)}$  and  $\tilde{\rho}_{m,n}^{(k,\tilde{k})} \in \{0, 1\}$ , then  $\tilde{\rho}_{m,n}^{(k,\tilde{k})} \leq \rho_{m,n}^{(k)}$ . Moreover, since  $\sum_{m=1}^{M(k)} \tilde{\rho}_{m,n}^{(k,\tilde{k})} \leq I_n^{(\tilde{k})}$  and  $I_n^{(\tilde{k})}, \tilde{\rho}_{m,n}^{(k,\tilde{k})} \in \{0, 1\}$ , then  $\sum_{m=1}^{M(k)} \tilde{\rho}_{m,n}^{(k,\tilde{k})} \leq 1$  which means that  $\tilde{\rho}_{m,n}^{(k,\tilde{k})} \leq I_n^{(\tilde{k})}$ . Hence,  $\tilde{\rho}_{m,n}^{(k,\tilde{k})} \in \mathcal{S}$  which means that  $\mathcal{S}' \subseteq \mathcal{S}$ .  $\square$

By replacing the term  $\rho_{m,n}^{(k)} \max_{\tilde{k}} I_n^{(\tilde{k})} \tilde{r}_{m,n}^{(k,\tilde{k})}$  in (4.13) with (4.14) and (4.17), we get the following binary linear optimization problem:

$$\begin{aligned} & \underset{\{\rho_{m,n}^{(k)}\}, \{\tilde{\rho}_{m,n}^{(k,\tilde{k})}\}, \{I_n^{(k)}\}}{\text{maximize}} && \sum_{k=1}^K \sum_{m=1}^{M^{(k)}} w_m^{(k)} \left( \rho_{m,n}^{(k)} r_{m,n}^{(k)} + \sum_{\tilde{k} \in \mathcal{K}^{(k)}} \tilde{\rho}_{m,n}^{(k,\tilde{k})} \tilde{r}_{m,n}^{(k,\tilde{k})} \right) \end{aligned} \quad (4.18a)$$

$$\text{subject to} \quad \sum_{m=1}^{M^{(k)}} \rho_{m,n}^{(k)} = 1 - I_n^{(k)}, \quad \forall k \in \mathcal{K}, \quad (4.18b)$$

$$\sum_{\tilde{k} \in \mathcal{K}^{(k)}} \tilde{\rho}_{m,n}^{(k,\tilde{k})} \leq \rho_{m,n}^{(k)}, \quad \forall k \in \mathcal{K}, m \in \mathcal{M}^{(k)}, \quad (4.18c)$$

$$\sum_{m=1}^{M^{(k)}} \tilde{\rho}_{m,n}^{(k,\tilde{k})} \leq I_n^{(\tilde{k})}, \quad \forall k \in \mathcal{K}, \tilde{k} \in \mathcal{K}^{(k)}, \quad (4.18d)$$

$$\rho_{m,n}^{(k)}, \tilde{\rho}_{m,n}^{(k,\tilde{k})}, I_n^{(k)} \in \{0, 1\}, \quad \forall k \in \mathcal{K}, \tilde{k} \in \mathcal{K}^{(k)}, m \in \mathcal{M}^{(k)}. \quad (4.18e)$$

We finally remark that many other equivalent binary optimization problems can be formulated, e.g., by replacing the term  $\rho_{m,n}^{(k)} \max_{\tilde{k}} I_n^{(\tilde{k})} \tilde{r}_{m,n}^{(k,\tilde{k})}$  in (4.13) with (4.14) and (4.15). However, different equivalents will have different relaxations. Unlike many other equivalent formulations, we found that the relaxed version of (4.18) has optimal solution that is provably close to binary as we will show in Section 4.4.3. Moreover, we show in Section 4.4.5 that the optimal solution of this relaxation can be obtained efficiently using network flow optimization tools.

### 4.4.3 Linear Optimization Relaxation

An upper bound on the optimum value of (4.18) can be obtained by solving the relaxed version of (4.18) which can be constructed by replacing (4.18e) with the following constraints

$$\rho_{m,n}^{(k)}, \tilde{\rho}_{m,n}^{(k,\tilde{k})}, I_n^{(k)} \in [0, 1], \quad \forall k \in \mathcal{K}, \tilde{k} \in \mathcal{K}^{(k)}, m \in \mathcal{M}^{(k)}. \quad (4.19)$$

In particular, let  $p_{\text{Binary}}^*$  denote the optimal value of (4.18), let  $p_{\text{Relaxed}}^*$  denote the optimal value of the relaxed version of (4.18), and let  $\hat{p}_{\text{Relaxed}}^*$  denote the value of the objective function evaluated at a rounded-solution of the relaxed problem, such that the rounded solution is a feasible binary solution of (4.18). Then, we have the following inequalities

$$p_{\text{Relaxed}}^* \geq p_{\text{Binary}}^* \geq \hat{p}_{\text{Relaxed}}^*, \quad (4.20)$$

where the first inequality follows from the fact that the feasible set of the relaxed version is always a superset of the original problem and the second equality follows directly from the optimality of  $p_{\text{Binary}}^*$ . We define the optimality gap,  $\Delta_{\text{Opt}}$ , in percentage as

$$\Delta_{\text{Opt}} = 100\% \cdot (p_{\text{Binary}}^* - \hat{p}_{\text{Relaxed}}^*) / p_{\text{Binary}}^* \leq 100\% \cdot (p_{\text{Relaxed}}^* - \hat{p}_{\text{Relaxed}}^*) / p_{\text{Relaxed}}^*, \quad (4.21)$$

where the inequality follows from (4.20). Thus, one can compute an estimate on the optimality gap in polynomial time by solving a linear optimization problem. As we will show in Section 4.6.1 through extensive simulations, by solving the relaxed problem and rounding the solution to the closest binary feasible solution, one can obtain a solution to (4.18) that is near-optimal, i.e., with small  $\Delta_{\text{Opt}}$ .

In addition to the optimality gap given in (4.21), which can be computed in polynomial time, we now provide a closed-form theoretical guarantee on the percentage of optimal relaxed variables that assume binary values. This theoretical guarantee is summarized in the following lemma.

**Lemma 6.** *The percentage of optimal variables of the relaxed version of (4.18) that assume binary values is greater than or equal to  $\frac{6(\bar{M}-1)}{7\bar{M}+1} \cdot 100\%$ , where  $\bar{M}$  is the average number of UTs per sector, i.e.,  $\bar{M} = \frac{1}{K} \sum_{k=1}^K M^{(k)}$ .*

*Proof.* We start by categorizing the constraints of the relaxed version of (4.18) into

two groups: non-bounding constraints, which are given by 4.18b, 4.18c, and 4.18d, and bounding constraints, which are given by 4.19. It is easy to see that the number of variables that assume optimal binary values is equal to the number of bounding inequality constraints that are active at the optimal, i.e., inequalities that are satisfied with equalities at the optimal.

We now note that the optimal solution of a linear program lies on one or more vertices. For a linear program with  $N_{var}$  variables, a vertex is characterized by  $N_{var}$  equalities or active inequalities. Let  $N_{non-bounding}^*$  denote the number of non-bounding equality or active inequality constraints at the optimum, and let  $N_{bounding}^*$  denote the number of active bounding inequality constraints at the optimum. Hence, we can write the following:

$$\begin{aligned}
N_{var} &= N_{bounding}^* + N_{non-bounding}^* \Rightarrow \\
N_{bounding}^* &= N_{var} - N_{non-bounding}^* \Rightarrow \\
N_{bounding}^* &\geq N_{var} - \sup(N_{non-bounding}^*) \Rightarrow \\
N_{bounding}^* &\geq 7 \sum_{k=1}^K M^{(k)} + K - (K + \sum_{k=1}^K M^{(k)} + 6K) = 6 \sum_{k=1}^K M^{(k)} - 6K,
\end{aligned} \tag{4.22}$$

where the supremum is found by assuming that all non-bounding constraints are active. Thus, the percentage of optimal variables of the relaxed version of (4.18) that assume binary values can be bounded as

$$P_{\text{binary}} \geq \frac{6 \sum_{k=1}^K M^{(k)} - 6K}{7 \sum_{k=1}^K M^{(k)} + K} = \frac{6(\bar{M} - 1)}{7\bar{M} + 1} \cdot 100\%, \tag{4.23}$$

where  $\bar{M}$  is the average number of UTs per sector, i.e.,  $\bar{M} = \frac{1}{K} \sum_{k=1}^K M^{(k)}$ .  $\square$

For example, if  $M^{(k)} = 10, \forall k$ , then using Lemma 6 we can deduce that the percentage of optimal variables of the relaxed version of (4.18) that assume binary

values is guaranteed to be greater than or equal to 76%.

Unfortunately, solving the relaxed version of (4.18) would require a central controller to be connected to all the BSs in order to solve a large linear optimization problem. Such a central controller is not supported in most contemporary cellular network standards, such as LTE, LTE-Advanced and IEEE 802.16m. Consequently, we seek a distributed optimization method to solve the relaxed version of problem (4.18).

#### 4.4.4 Primal Decomposition

The relaxed version of (4.18) has a special separable structure. In particular, for any set of fixed  $\{I_n^{(k)}, \forall k \in \mathcal{K}\}$ , it can be separated into  $K$  optimization problems, each can be solved separately in each sector. In this section, we show how this structure is exploited to develop a distributed algorithm based on the primal-decomposition method [90, pp. 3–5].

To exploit the separable structure, let  $\phi^{(k)}(I_n^{(1)}, \dots, I_n^{(K)})$  denote the optimal value of the following optimization problem for given  $\{I_n^{(1)}, \dots, I_n^{(K)}\}$ :

$$\underset{\{\rho_{m,n}^{(k)}, \{\tilde{\rho}_{m,n}^{(k,\tilde{k})}\}}}{\text{maximize}} \quad \sum_{m=1}^{M^{(k)}} w_m^{(k)} \left( \rho_{m,n}^{(k)} r_{m,n}^{(k)} + \sum_{\tilde{k} \in \mathcal{K}^{(k)}} \tilde{\rho}_{m,n}^{(k,\tilde{k})} \tilde{r}_{m,n}^{(k,\tilde{k})} \right), \quad (4.24a)$$

$$\text{subject to} \quad \sum_{m=1}^{M^{(k)}} \rho_{m,n}^{(k)} = 1 - I_n^{(k)}, \quad (4.24b)$$

$$\sum_{\tilde{k} \in \mathcal{K}^{(k)}} \tilde{\rho}_{m,n}^{(k,\tilde{k})} \leq \rho_{m,n}^{(k)}, \quad \forall m \in \mathcal{M}^{(k)}, \quad (4.24c)$$

$$\sum_{m=1}^{M^{(k)}} \tilde{\rho}_{m,n}^{(k,\tilde{k})} \leq I_n^{(\tilde{k})}, \quad \forall \tilde{k} \in \mathcal{K}^{(k)}, \quad (4.24d)$$

$$\rho_{m,n}^{(k)}, \tilde{\rho}_{m,n}^{(k,\tilde{k})} \in [0, 1], \quad \forall m \in \mathcal{M}^{(k)}, \forall \tilde{k} \in \mathcal{K}^{(k)}. \quad (4.24e)$$

For reasons that will become apparent, we call (4.24) subproblem  $k$ . Using (4.24),

the relaxed version of (4.18) is equivalent to

$$\text{maximize}_{\{I_n^{(k)}\}} \sum_{k=1}^K \phi^{(k)}(I_n^{(1)}, \dots, I_n^{(K)}) \quad (4.25a)$$

$$I_n^{(k)} \in [0, 1], \quad \forall k \in \mathcal{K}. \quad (4.25b)$$

We call (4.25) the master problem. Therefore, the relaxed version of (4.18) has been decomposed into a master problem, given by (4.25), and  $K$  subproblems, each is given by (4.24) and can be solved separately in each sector.

The master problem given in (4.25) can be solved iteratively using the projected-subgradient method [91, p. 16]. In each iteration,  $K$  subproblems are solved in each sector in order to evaluate  $\phi^{(k)}(I_n^{(1)}, \dots, I_n^{(K)})$ ,  $\forall k \in \mathcal{K}$ , and subgradients  $[\Lambda_n^{*(1)}, \dots, \Lambda_n^{*(K)}] \in \partial \sum_{k=1}^K \phi^{(k)}(I_n^{(1)}, \dots, I_n^{(K)})$ , where  $\partial f(x)$  is the subdifferential of  $f(\cdot)$  evaluated at  $x$ . A subgradient  $\Lambda_n^{*(k)}$  is calculated as [91]

$$\Lambda_n^{*(k)} := -\lambda_n^{*(k)} + \sum_{\tilde{k} \in \mathcal{K}^{(k)}} \lambda_n^{*(\tilde{k}, k)}, \quad (4.26)$$

where  $\lambda_n^{*(k)}$  is an optimum Lagrange multiplier (dual variable) corresponding to constraint (4.24b) and  $\lambda_n^{*(k, \tilde{k})}$ ,  $\tilde{k} \in \mathcal{K}^{(k)}$ , are optimum Lagrange multipliers corresponding to constraints (4.24d). In order for each sector  $k$  to calculate  $\Lambda_n^{*(k)}$ , it requires the knowledge of  $\lambda_n^{*(k)}$ , which can be obtained locally by solving (4.24), and  $\lambda_n^{*(\tilde{k}, k)}$ ,  $\tilde{k} \in \mathcal{K}^{(k)}$ , which can be exchanged from the neighboring sectors. In other words, each sector  $k$  sends  $\lambda_n^{*(k, \tilde{k})}$  for all  $\tilde{k}$  sectors that are in the neighborhood of sector  $k$ , for all  $n$ . The master algorithm then updates its variables as

$$I_n^{(k)} := I_n^{(k)} + \delta \Lambda_n^{*(k)}, \quad \forall k \in \mathcal{K}, \quad (4.27)$$

where  $\delta$  is the step-size which can be chosen using any of the standard methods given

in [91, pp. 3–4]. Then, each  $I_n^{(k)}$  is projected into the feasible set of  $[0, 1]$  as follows

$$I_n^{(k)} := \begin{cases} 0, & I_n^{(k)} \leq 0, \\ I_n^{(k)}, & 0 < I_n^{(k)} < 1, \\ 1, & I_n^{(k)} \geq 1. \end{cases} \quad (4.28)$$

Each sector  $k$  exchanges  $I_n^{(k)}$  with its neighbors and the process is repeated for  $N_{\text{iter}}$  iterations. Then, each  $I_n^{(k)}$  is rounded to the nearest binary value which is denoted by  $I_n^{\star(k)}$ . Once  $\{I_n^{\star(k)}\}$  are determined, local scheduling decisions can be done in each sector  $k$  separately as follows. For every  $m \in \mathcal{M}^{(k)}, n \in \mathcal{N}$ ,  $R_{m,n}^{(k)}$  is calculated using (4.2), (4.1), and  $I_n^{\star(k)}$ . Finally, to ensure feasibility of the resulting solution to problem (4.18), the scheduling decision variables are calculated as

$$\rho_{m,n}^{\star(k)} = \begin{cases} 1, & m = \arg \max_m w_m^{(k)} R_{m,n}^{(k)} \text{ and } I_n^{\star(k)} = 0, \\ 0, & m \neq \arg \max_m w_m^{(k)} R_{m,n}^{(k)} \text{ or } I_n^{\star(k)} = 1. \end{cases} \quad (4.29)$$

The sub-gradient algorithm is guaranteed to converge to the optimum solution of the relaxed version of problem (4.18) as  $N_{\text{iter}} \rightarrow \infty$  if  $\delta$  is chosen properly [91, p. 6]. In this chapter, we choose  $\delta$  to be square summable but not summable by setting  $\delta = c/p$ , where  $c > 0$  is a constant and  $p$  is the iteration index. This choice guarantees convergence to the optimal solution as  $N_{\text{iter}} \rightarrow \infty$  [91, p. 6]. The constant  $c$  is a parameter that can be tuned to improve the convergence time. Through extensive simulations, we found that a good choice of the parameter  $c$  is mainly dependant on the choice of the desired utility. Thus, for a given utility, the parameter  $c$  can be tuned offline only once before executing the algorithm. In practise, however, the algorithm must terminate after a finite number of iterations which raises the following

question: How different is the value obtained using finite iterations as compared to the true optimum obtained by solving (4.18)? We address this question in Section 4.6 and show that few iterations are sufficient to achieve near-optimality.

Clearly, the proposed algorithm relies heavily on solving the subproblem given by (4.24). Hence, it is imperative to solve (4.24) as efficiently as possible. Interestingly, the subproblem given by (4.24) has a special network flow structure which can be exploited to devise efficient algorithms to solve it, as explained in the following section.

#### **4.4.5 Transforming (4.24) into an Equivalent Minimum Cost Network Flow Problem**

The optimization problem (4.24) is a linear optimization problem which can be solved using generic simplex or interior-point methods. Nevertheless, we show in this section that (4.24) has a special network structure which makes it amenable to powerful network flow optimization methods that surpass conventional simplex and interior-point methods. In particular, we show that (4.24) can be converted into an equivalent minimum cost network flow optimization problem.

A minimum cost network flow optimization problem is defined as finding a least cost way of sending certain amount of flow over a network that is specified by a directed graph of  $v$  vertices and  $e$  edges. Such a problem has  $e$  variables, which represent the amount of flow on each arc, and  $v$  linear equality constraints, which represent the mass-balance in each vertex, such that every variable appears in exactly two constraints: one with a coefficient of  $+1$  and one with a coefficient of  $-1$  [4, p. 5]. In addition to the mass-balance constraints, constraints on the lower and upper bounds on the amount of flow on each arch are also specified. The objective function is a weighted sum of the flows in each arc, where the weight is the cost per unit

flow on that arc. Thanks to the network structure of these problems, highly efficient combinatorial algorithms exist to solve such problems in strongly polynomial time [4, Chapter 10]. These algorithms perform much faster than a generic linear optimization solver based on conventional simplex or interior-point methods. For example, the enhanced capacity scaling algorithm can solve a minimum cost network flow problem with  $v$  vertices and  $e$  edges in  $O(e \log v(e + v \log v))$  [4, p. 395].

By examining (4.24), we see that it is not a minimum cost network flow optimization problem (in its current form) as it does not have mass-balance constraints. Nevertheless, if we multiply both sides of constraint (4.24d) with  $-1$ , we obtain a linear optimization problem with the following properties. Each variable appears in at most one constraint with a coefficient of  $+1$  and at most one constraint with a coefficient of  $-1$ . According to Theorem 9.9 in [4, p. 315], a linear optimization problem with such a structure can be transformed into an equivalent minimum cost network flow optimization problem. To perform such transformation, we introduce the slack variables  $\{s_m \geq 0, \forall m \in \mathcal{M}^{(k)}\}$  and surplus variables  $\{\tilde{s}^{(\tilde{k})} \geq 0, \forall \tilde{k} \in \mathcal{K}^{(k)}\}$  to convert the inequality constraints (4.24c) and (4.24d), respectively, into equality constraints. To obtain the mass-balance constraints, we also introduce a redundant constraint by summing constraints (4.30b)–(4.30d). In addition, we convert the maximization into minimization by negating the objective function. Incorporating these transformation into (4.24), we get the following minimum cost network flow optimization problem:

$$\begin{aligned} & \underset{\{\rho_{m,n}^{(k)}\}, \{\tilde{\rho}_{m,n}^{(k,\tilde{k})}\}, \{s_m\}, \{\tilde{s}^{(\tilde{k})}\}}{\text{minimize}} && - \sum_{m=1}^{M^{(k)}} w_m^{(k)} \left( \rho_{m,n}^{(k)} r_{m,n}^{(k)} + \sum_{\tilde{k} \in \mathcal{K}^{(k)}} \tilde{\rho}_{m,n}^{(k,\tilde{k})} \tilde{r}_{m,n}^{(k,\tilde{k})} \right) \end{aligned} \quad (4.30a)$$

$$\text{subject to} \quad \sum_{m=1}^{M^{(k)}} \rho_{m,n}^{(k)} = 1 - I_n^{(k)}, \quad (4.30b)$$

$$\sum_{\tilde{k} \in \mathcal{K}^{(k)}} \tilde{\rho}_{m,n}^{(k,\tilde{k})} - \rho_{m,n}^{(k)} + s_m = 0, \quad \forall m \in \mathcal{M}^{(k)}, \quad (4.30c)$$

$$- \sum_{m=1}^{M^{(k)}} \tilde{\rho}_{m,n}^{(k,\tilde{k})} - \tilde{s}^{(\tilde{k})} = -I_n^{(\tilde{k})}, \quad \forall \tilde{k} \in \mathcal{K}^{(k)}, \quad (4.30d)$$

$$- \sum_{m=1}^{M^{(k)}} s_m + \sum_{\tilde{k} \in \mathcal{K}^{(k)}} \tilde{s}^{(\tilde{k})} = -1 + I_n^{(k)} + \sum_{\tilde{k} \in \mathcal{K}^{(k)}} I_n^{(\tilde{k})}, \quad (4.30e)$$

$$\rho_{m,n}^{(k)}, \tilde{\rho}_{m,n}^{(k,\tilde{k})}, s_m, \tilde{s}^{(\tilde{k})} \in [0, 1], \quad \forall m \in \mathcal{M}^{(k)}, \tilde{k} \in \mathcal{K}^{(k)}. \quad (4.30f)$$

To see that (4.30) is indeed a minimum cost network flow optimization problem, we note that the objective function is linear and the equality constraints (4.30b)-(4.30e) represent mass-balance constraints because each variable appears in two constraints: one with a coefficient of +1 and one with a coefficient of  $-1$  [4, p. 5].

To illustrate graphically the minimum cost network flow structure of (4.30), we consider a simple example of 4 sectors and 2 UTs per sector. For simplicity, we focus on the optimization problem solved by sector 1. We also consider a single RB and thus drop the RB index  $n$ . For the considered example, the optimization problem (4.30) for sector 1 can be written as

$$\begin{aligned} \text{minimize}_{\substack{\{\rho_m^{(1)}\}, \{\tilde{\rho}_m^{(1,\tilde{k})}\}, \\ \{s_m\}, \{\tilde{s}^{(\tilde{k})}\}}} & - \sum_{m=1}^2 w_m^{(1)} \left( \rho_m^{(1)} r_m^{(1)} + \sum_{\tilde{k}=2}^4 \tilde{\rho}_m^{(1,\tilde{k})} \tilde{r}_m^{(k,\tilde{k})} \right) \end{aligned} \quad (4.31a)$$

$$\text{subject to} \quad \sum_{m=1}^2 \rho_m^{(1)} = 1 - I^{(1)}, \quad (4.31b)$$

$$\sum_{\tilde{k}=2}^4 \tilde{\rho}_1^{(1,\tilde{k})} - \rho_1^{(1)} + s_1 = 0, \quad (4.31c)$$

$$\sum_{\tilde{k}=2}^4 \tilde{\rho}_2^{(1,\tilde{k})} - \rho_2^{(1)} + s_2 = 0, \quad (4.31d)$$

$$- \sum_{m=1}^2 \tilde{\rho}_m^{(1,2)} - \tilde{s}^{(2)} = -I^{(2)}, \quad (4.31e)$$

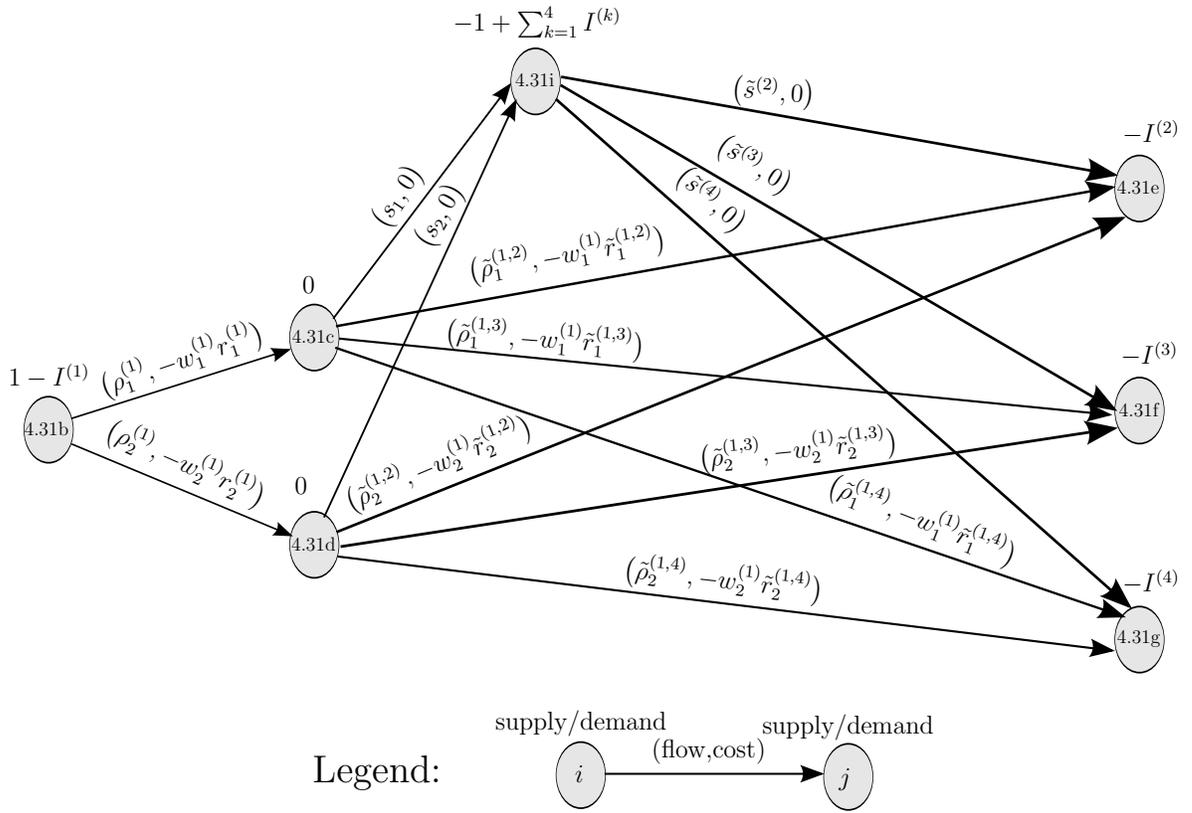
$$- \sum_{m=1}^2 \tilde{\rho}_m^{(1,3)} - \tilde{s}^{(3)} = -I^{(3)}, \quad (4.31f)$$

$$- \sum_{m=1}^2 \tilde{\rho}_m^{(1,4)} - \tilde{s}^{(4)} = -I^{(4)}, \quad (4.31g)$$

$$- \sum_{m=1}^2 s_m + \sum_{\tilde{k}=2}^4 \tilde{s}^{(\tilde{k})} = -1 + I^{(1)} + \sum_{\tilde{k}=2}^4 I^{(\tilde{k})}, \quad (4.31h)$$

$$\rho_m^{(1)}, \tilde{\rho}_m^{(1,\tilde{k})}, s_m, \tilde{s}^{(\tilde{k})} \in [0, 1], \quad \forall m \in \{1, 2\}, \tilde{k} \in \{2, 3, 4\}. \quad (4.31i)$$

In Fig. 4.5, we provide an equivalent graph representation of the minimum cost network flow problem given by (4.31). In this graph, each vertex represents a constraint in (4.31) and it is labeled according to the label of its corresponding constraint. A vertex is called a supply vertex if its total outflow (i.e., flow emanating from it) is greater than its total inflow (i.e., flows entering it), a demand vertex if its total outflow is less than its total inflow, and it is called a transshipment vertex if its total outflow is equal to its total inflow. Hence, vertex (4.31b) is a supply vertex, vertices (4.31e)–(4.31g) are demand vertices, vertices (4.31c)–(4.31d) are transshipment vertices, and vertex (4.31i) can be any of the above mentioned types depending on



**Figure 4.5:** Equivalent graphical representation of the minimum cost network flow problem given by (4.31). Vertices are labeled according to the label of their corresponding constraints. All flows are restricted to be between 0 and 1.

the values of  $\{I^{(k)}\}$ . Clearly, the constraints correspond to the vertices in this graph are mass-balance constraints as the total outflow minus the total inflow is equal to the supply or demand of these vertices. Similar to the vertices, each edge is labeled with a flow, which represents the corresponding variable, and a cost per unit flow, which represent the corresponding coefficient in the objective function. The flow in each edge is restricted to be in the range  $[0, 1]$ . Based on the above, we conclude that the optimization problem in (4.31) is equivalent to finding the least cost way of sending  $1 - I^{(1)}$  flow units from the supply vertex (4.31b) to the demand vertices (4.31e)–(4.31g) in the graph shown in Fig. 4.5.

#### 4.4.6 Pseudocode

A pseudocode of the proposed algorithm to be executed in every sector  $k$  is given below.

---

**Algorithm 4** Proposed ICIC algorithm to be executed in every sector  $k$

---

**Input:**  $H_{m,n}^{(k,\tilde{k})}$  and  $w_m^{(k)}$ ,  $\forall \tilde{k} \in \mathcal{K}^{(k)}, m \in \mathcal{M}, n \in \mathcal{N}$

**Output:**  $\rho_{m,n}^{*(k)}, I_n^{*(k)}$ ,  $\forall m \in \mathcal{M}, n \in \mathcal{N}$

- 1: Initialize  $I_n^{(k)}$ ,  $\forall n \in \mathcal{N}$
  - 2: **for**  $p = 1$  to  $N_{\text{iter}}$  **do**
  - 3:   **Preprocessing:** Obtain  $r_{m,n}^{(k)}$  and  $\tilde{r}_{m,n}^{(k,\tilde{k})}$  using (4.10) and (4.11),  $\forall \tilde{k} \in \mathcal{K}^{(k)}, m \in \mathcal{M}, n \in \mathcal{N}$ .
  - 4:   **Solve a subproblem:** Obtain  $\rho_{m,n}^{*(k)}, \tilde{\rho}_{m,n}^{*(k,\tilde{k})}, \lambda_n^{*(k)}$ , and  $\lambda_n^{*(k,\tilde{k})}$ ,  $\forall \tilde{k} \in \mathcal{K}^{(k)}, n \in \mathcal{N}$ , by solving the minimum cost network flow optimization problem (4.30),  $\forall n \in \mathcal{N}$ .
  - 5:   **Exchange sub-gradients:** Send  $\lambda_n^{*(k,\tilde{k})}$  to sectors  $\tilde{k} \in \mathcal{K}^{(k)}$ ,  $\forall n \in \mathcal{N}$ .
  - 6:   **Update**  $I_n^{(k)}$ ,  $\forall n \in \mathcal{N}$ :
  - 7:      $\Lambda_n^{*(k)} := -\lambda_n^{*(k)} + \sum_{\tilde{k} \in \mathcal{K}^{(k)}} \lambda_n^{*(\tilde{k},k)}$ .
  - 8:    **Subgradient step:**  $I_n^{(k)} := I_n^{(k)} + \delta \Lambda_n^{*(k)}$
  - 9:    **Projection:** Project  $I_n^{(k)}$  into the feasible set (4.28).
  - 10:   **Exchange**  $I_n^{(k)}$ : Send  $I_n^{(k)}$  to sectors  $\tilde{k} \in \mathcal{K}^{(k)}$ ,  $\forall n \in \mathcal{N}$ .
  - 11: **end for**
  - 12: **Round the solution:**  $I_n^{*(k)} := \lfloor I_n^{(k)} + 0.5 \rfloor$ ,  $\forall n \in \mathcal{N}$ .
  - 13: **Local scheduling decisions:**
  - 14:   Obtain  $R_{m,n}^{(k)}$  by substituting  $H_{m,n}^{(k,\tilde{k})}$  and  $I_n^{*(k)}$  in (4.1) and (4.2),  $\forall m \in \mathcal{M}, n \in \mathcal{N}$ .
  - 15:   Obtain  $\rho_{m,n}^{*(k)}$  by substituting  $w_m^{(k)}$  and  $R_{m,n}^{(k)}$  in (4.29),  $\forall m \in \mathcal{M}, n \in \mathcal{N}$ .
- 

#### 4.4.7 Intuitive Interpretation

The proposed algorithm, whose pseudocode is given in Algorithm 4, has the following intuitive interpretation. Each sector  $k$  estimates the benefit of shutting its own power on RB  $n$  to neighboring sectors by  $\sum_{\tilde{k} \in \mathcal{K}^{(k)}} \lambda_n^{*(\tilde{k},k)}$  and the benefit of using RB  $n$  in sector  $k$  by  $\lambda_n^{*(k)}$ ; if the former is more (less) than the latter, then sector  $k$  increases (decreases) its soft decision on  $I_n^{(k)}$  proportionally to the difference of these benefits, and possibly set it to 1 (0).

#### 4.4.8 Complexity Analysis of the Proposed Algorithm

In the following, we show that the worst-case time complexity of the proposed algorithm implemented in each sector  $k$  is polynomial. We start by analyzing the complexity of solving the minimum cost network flow problem given by (4.30). A graph that is equivalent to (4.30) is represented by  $8+M^{(k)}$  nodes, which is equal to the number of mass-balance constrains, and  $8M^{(k)}+6$  arcs, which is equal to the number of variables. Thus, (4.30) can be solved using the enhanced capacity scaling algorithm in  $O((M^{(k)} \log M^{(k)})^2)$  [4, p. 395]. Since (4.30) is solved for each RB in  $N_{\text{iter}}$  iterations, then the complexity of the proposed algorithm is  $O(N_{\text{iter}}N(M^{(k)} \log M^{(k)})^2)$ . Since  $N_{\text{iter}}$  is a fixed constant, the complexity of the algorithm in sector  $k$  as a function of  $M^{(k)}$  and  $N$  is  $O(N(M^{(k)} \log M^{(k)})^2)$ .

Having a polynomial algorithm is crucial for the scalability of the network. Indeed, we observed through extensive simulations that the average simulation time grows as  $MN$ , i.e., linearly in the number of RBs and the number of UTs.

To reduce the information exchange between sectors, low complexity variants of the proposed algorithm can be implemented. One such implementation is to execute the algorithm every  $\tau$  sub-frames. Each sector has time to implement the algorithm and calculate the  $\{I_n^{(k)}\}$  values during  $\tau$  sub-frames and exchange them with neighboring BSs. The design parameter  $\tau$  can be adjusted to suit the practical limitations. Also, a pipelined implementation is possible, where the algorithm is executed every sub-frame but with a delayed decision of  $\tilde{\tau}$  sub-frames.

We now quantify the rate of message exchange required between any two neighboring sectors, in order to execute the proposed algorithm. For each iteration required by the proposed algorithm, each sector sends to its neighboring sector  $N$  subgradients values and  $N$   $I$  values. Thus, assuming  $L_s$  bits is used to quantize the subgradients and  $I$  values, then it is not difficult to see that the rate of message exchange (in b/s)

is given by

$$R_{\text{exchange}} = \frac{2N_{\text{iter}}NL_q}{\tau \text{1ms}} \text{ (b/s)}. \quad (4.32)$$

For example, for  $N_{\text{iter}} = 5$  iterations,  $N = 50$  RBs,  $L_q = 8$  bits, and  $\tau = 10$  sub-frames,  $R_{\text{exchange}} = 400$  Kb/s. Such data rate can be realized in practice through either high-speed wired links or high-speed LOS wireless links since the locations of the sectors are fixed.

## 4.5 Baseline Schemes

In this chapter, the following are used as the baseline schemes: reuse-1, reuse-3, PFR [42], dynamic FFR [46], and optimum FFR [55]. While reuse-1, reuse-3, and PFR are static schemes, dynamic and optimum FFR are implemented as dynamic schemes.

In PFR, RBs are divided such as 30 RBs (inner band) are used in all sectors (reuse-1) while 20 RBs (outer band) are shared among sites in a classical reuse-3 pattern. More details about this algorithm can be found in [42].

While dynamic FFR and optimum FFR were designed to maximize the sum-rate, they can be easily modified to accommodate maximizing weighted sum-rates, and they work with differentiable and non-differentiable AMC functions, which make them good baseline schemes to be compared with the proposed scheme. For fair comparison, we convert the constraint on instantaneous minimum rate (used in dynamic and optimum FFR) to a constraint on average minimum rate, which can be implemented by updating the weights in every sub-frame according to the procedure given in [33]. We finally remark that both dynamic FFR and optimum FFR are centralized schemes.

In dynamic FFR, a centralized controller is used to determine which RBs belong to inner band (in reuse-1 pattern) and which RBs belong to outer bands (in reuse-3 pattern). The decision is made to maximize the total utility, assuming each RB is

used by all the users in the region where this RB can be used, and summed for all the sectors. This assumption is critical to the development of the algorithm. Then, scheduling is done locally by each BS. More details about this algorithm can be found in [46].

In optimum FFR, the number of RBs used in inner and outer bands, and the reuse factor of the outer band are determined optimally using a centralized controller, without considering channel fading. Without channel fading, all RBs seen by a particular user have the same SINR. This makes the optimization problem tractable. However, if channel fading is considered, then the computational complexity for optimum FFR is exponential. More details about this algorithm can be found in [55].

It is common for FFR schemes (e.g., PFR and optimum FFR) to divide users into two classes: inner (cell-center) users and outer (cell-edge) users, based on SINR or distance from BS. Inner users are restricted to use inner band while outer users are restricted to use outer band. We found such restriction degrades the performance of FFR and as such, this restriction is removed to realize the full potential of FFR schemes. This restriction is already removed from dynamic FFR [46] for the same reason.

## 4.6 Simulation Results

In this Section, we present the simulation results of the proposed scheme as well as the baseline schemes provided in Section 4.5. The simulation results were obtained through Monte Carlo simulations that are carried over  $10^3$  sub-frames and averaged over 10 independent drops. The simulation parameters and assumptions are based on the IMT-Advanced UMa scenario [34] which were explained in detail in Section 3.6.1. Validation of the simulator was presented in Section 3.6.2.

### 4.6.1 Optimality Gaps

To develop an efficient algorithm to solve the difficult optimization problem given by (4.13) in a distributed manner, two sources of sub-optimality were introduced, namely, relaxing the integer constraints and solving the master optimization problem in finite iterations. To understand the effect of these sources of sub-optimality, we present in Table 4.1.a the mean and standard deviation of the upper-bound on the optimality gap (4.21) for different number of iterations. For each number of iterations, we simulated a total of 22,000 instances of the optimization problems for four IMT-Advanced scenarios and different numbers of UTs [92]. In the first sub-frame, a random initial point is used. In the subsequent sub-frames, the optimal solution of the previous frame is used as an initial point. It is clear from Table 4.1.a that the proposed algorithm converges quickly to a near-optimum solution. Hence, for the rest of the simulations, the number of iterations is set to be 5, since this value results in a small optimality gap.

The numbers in Table 4.1.a illustrate that our algorithm can solve the bound optimization problem (4.13) in a near-optimum manner. A natural question to ask is how different is the optimal value obtained by solving the bound optimization as compared to the optimal value obtained by solving the original strongly NP-hard problem (4.7) using exhaustive search? Due to the exponential computational complexity of exhaustive search, simulating a system of 57 sectors is not feasible. As a result, we only show the optimality gap for a system of 12 sectors in Table 4.1.b, for proportional-fair scheduling, i.e.,  $\alpha$ -fair scheduling with  $\alpha = 1$ , for the cases of 10, 20, and 30 UTs per sector. As we can see, the proposed scheme can achieve, on average, about 96% of the optimum value achieved using exhaustive search, if it is executed once (5 iterations). One can also reduce the optimality gap further by executing the algorithm more than once. That is, after finding the optimum  $\{I_n^{(k)}\}$

**Table 4.1:** Mean and standard deviation of the optimality gap (%)

a. Compared to the optimal value of the bound optimization (4.13)

Number of iterations	Mean (%)	Standard deviation (%)
0	43.62	1.16
1	4.61	4.19
2	2.87	3.39
5	1.43	1.43
10	0.95	0.88

b. Compared to the optimal value of the original problem (4.7)

Number of runs	Mean (%)	Standard deviation (%)
1	3.8	1.7
2	2.4	0.8

in one run, the algorithm sets  $H_{m,n}^{(\tilde{k},k)} = 0$ ,  $\forall m, n, k, \tilde{k} | I_n^{(k)} = 1$ , and executes the algorithm again. In this case, one can achieve about 97.6% of the optimum value achieved using exhaustive search, i.e., an incremental gain, at the expense of more computational complexity. This suggests that the bound optimization is indeed a good method to achieve near-optimality and shows that most of the gain is already captured by executing the algorithm only once.

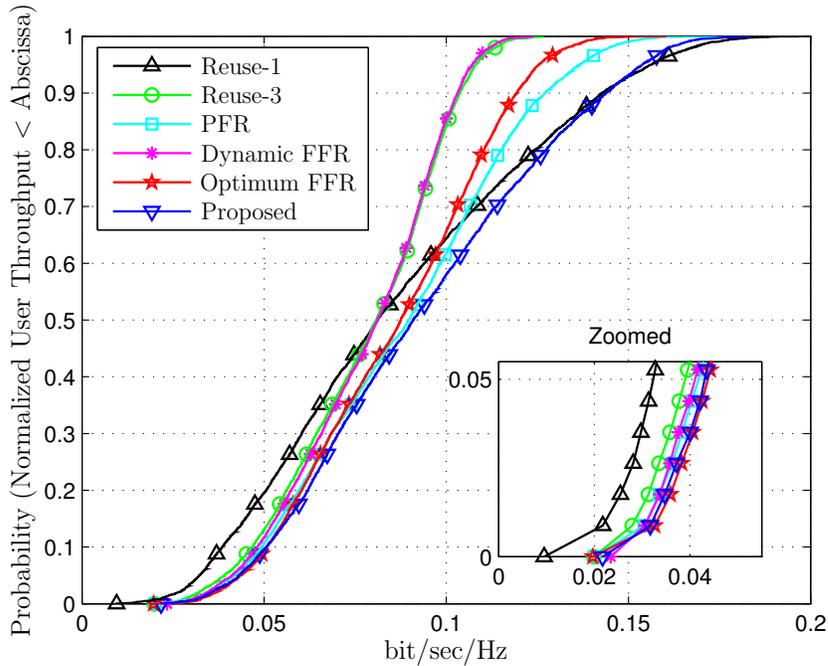
We conclude the discussion on optimality gap by noting that, since the original problem is strongly NP-hard (cf. [79]), there is no polynomial-time algorithm, including the proposed algorithm, as of yet that can provide guarantee on the optimality gap for the worst case scenario. However, we note that the optimality gap for the proposed algorithm is small for the case when each UTs is mainly interfered by at most one interferer. This is the case for BS deployment with highly directional antennas.

## 4.6.2 Comparing the Proposed Scheme with the Baseline Schemes

In this section, we compare the performance of the proposed scheme with the baseline schemes presented in Section 4.5. Similar to Section 3.6.3, we present the comparison results in terms of normalized throughputs measured in (bit/sec/Hz), where the term “normalized throughput” is defined as the long-term average rate divided by the total bandwidth (10 MHz). Moreover, normalized efficiency is defined as the normalized aggregate sector throughput which is obtained by summing the normalized user throughput of all users in the network and dividing the result by the number of sectors.

### Throughput CDF Comparison

In Fig. 4.6, we show the CDF of the normalized time-average UT throughput for all schemes. In all schemes,  $\alpha$ -fair scheduler is used with fairness exponent of  $\alpha = 2$  (cf. Section 4.3). To facilitate the comparison, we define the normalized cell-edge and cell-center user throughputs as the 5<sup>th</sup> and the 95<sup>th</sup> percentiles of the normalized user throughputs, respectively. It is clear from the figure that reuse-1 has the worst cell-edge performance (0.0323 (bit/sec/Hz)) as compared to the other three schemes, due to the excessive interference experienced at the cell-edge. Reuse-3, PFR, dynamic FFR, and optimum FFR, and the proposed scheme achieve normalized cell-edge user throughputs of 0.0391, 0.0417, 0.0412, 0.0435, and 0.0430 (bit/sec/Hz), respectively. However, reuse-3, PFR, dynamic FFR, and optimum FFR improve the cell-edge performance at the expense of reducing the overall throughput, especially for UTs close to the cell-center. For example, the 95<sup>th</sup> percentile achieved by reuse-3, PFR, dynamic FFR, and optimum FFR are 0.108, 0.136, 0.107, and 0.125 (bit/sec/Hz), respectively, as compared to 0.156 and 0.154 (bit/sec/Hz) achieved by reuse-1 and

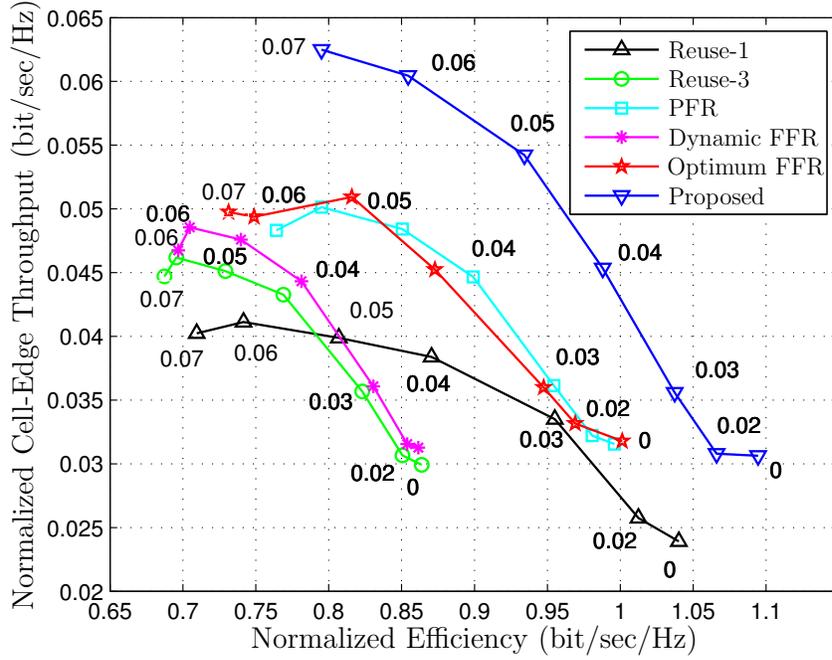


**Figure 4.6:** CDF of the normalized user throughput of all UTs in the network for  $\alpha = 2$  and  $\bar{R}_{\min} = 0$ .

the proposed scheme, respectively. Interestingly, the proposed scheme combines the advantages of all schemes, as it provides high cell-edge and cell-center throughputs simultaneously. Indeed, the gain achieved by the proposed scheme increases for higher fairness exponent,  $\alpha$ , or higher  $\bar{R}_{\min}$ , as we will see shortly.

### Efficiency-Cell-edge Tradeoff Comparison

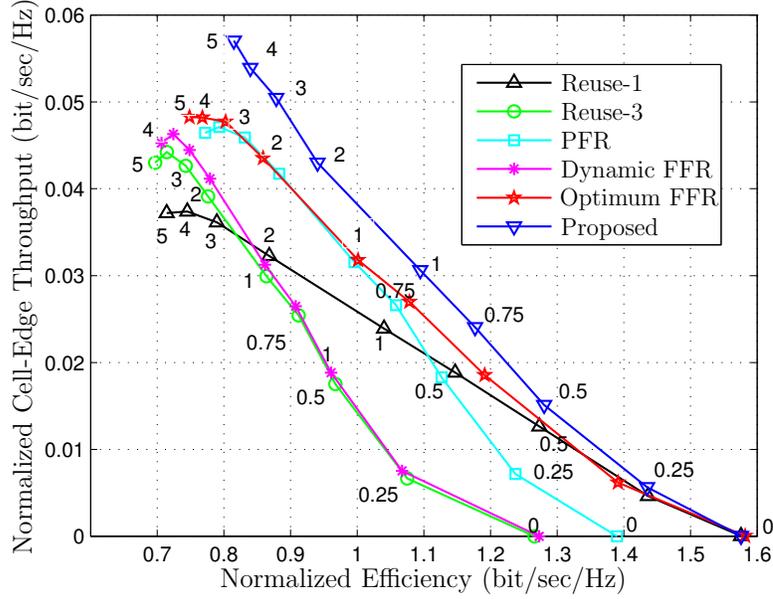
To further examine the performance of the different schemes, we show in Fig. 4.7 the normalized cell-edge user throughput and the normalized efficiency for all schemes and for different minimum average rate requirements  $\bar{R}_{\min}$ , assuming proportional-fair scheduler, i.e.,  $\alpha = 1$ . As expected, the general trend for all schemes is that as  $\bar{R}_{\min}$  increases, the normalized efficiency decreases and the normalized cell-edge user throughput increases. For high  $\bar{R}_{\min}$ , we observe that reuse-3, PFR, optimum FFR, dynamic FFR, and the proposed scheme have significantly higher normalized cell-edge



**Figure 4.7:** Normalized cell-edge throughput versus normalized efficiency, for  $\alpha = 1$  and  $\bar{R}_{\min} \in \{0, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07\}$ .

throughput than reuse-1. On the other hand, for small  $\bar{R}_{\min}$  it is clear that reuse-3, PFR, dynamic FFR, and optimum FFR incur significant loss in the normalized efficiency as compared to reuse-1. Interestingly, the proposed scheme performs very well in both the normalized cell-edge and the normalized efficiency as compared to all other schemes for all values of  $\bar{R}_{\min}$ . We also plot similar tradeoff curves in Fig. 4.8 by varying the fairness exponent  $\alpha$  and fixing  $\bar{R}_{\min} = 0$ , similar to [7]. Again, the proposed scheme outperforms other schemes in both the normalized cell-edge and the normalized efficiency.

In Figs 4.9 and 4.10, we take a closer look at the gains achieved by the different schemes as compared to reuse-1. In Fig. 4.9, we plot the gains in normalized cell-edge throughput achieved for a given normalized efficiency. The gain in normalized cell-edge throughput for a particular scheme for a given normalized efficiency  $x$  is given by  $\text{Gain}_{\text{scheme}}(x) = \frac{\text{CellEdge}_{\text{scheme}}(x) - \text{CellEdge}_{\text{reuse1}}(x)}{\text{CellEdge}_{\text{reuse1}}(x)} \cdot 100\%$ , where  $\text{CellEdge}_{\text{scheme}}(x)$  is the

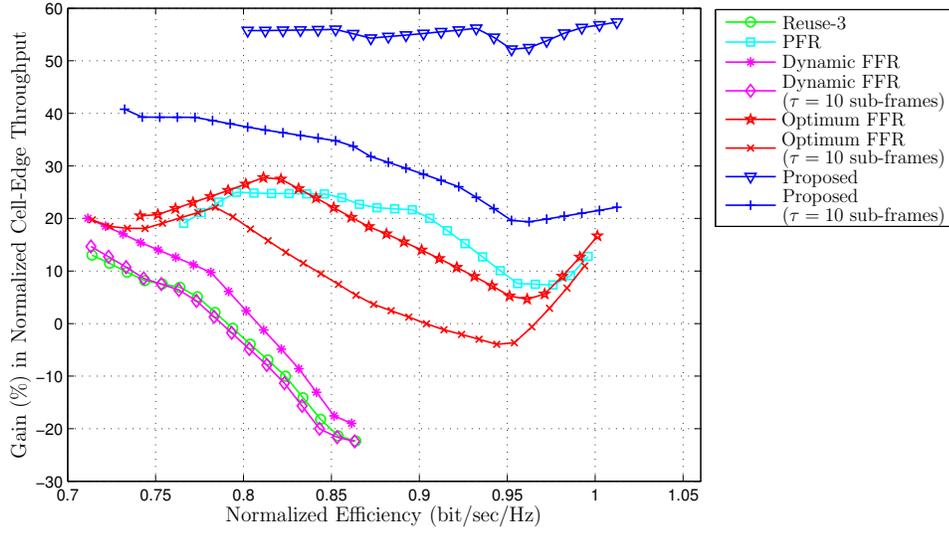


**Figure 4.8:** Normalized cell-edge throughput versus normalized efficiency for different schemes, for  $\bar{R}_{\min} = 0$  and  $\alpha \in \{0, 0.25, 0.50, 0.75, 1, 2, 3, 4, 5\}$ .

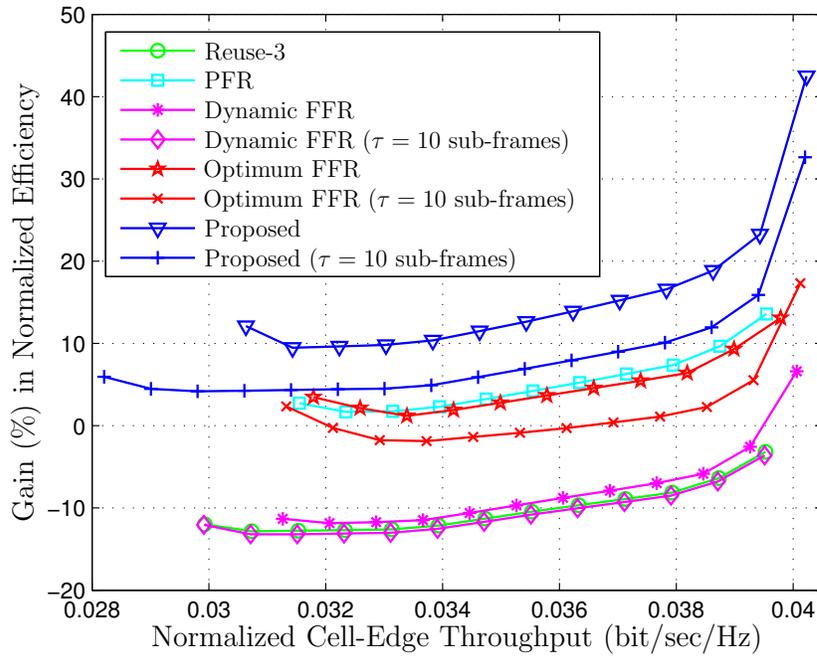
normalized cell-edge throughput achieved by a particular scheme at a normalized efficiency of  $x$ , which can be obtained from the tradeoff curves given in Fig. 4.7. For a wide range of normalized efficiencies, the proposed scheme achieves large gains (50% to 60%). Dynamic FFR, optimum FFR, and the proposed scheme lose some of the gain if it is executed every 10 sub-frames (as expected for any dynamic scheme); however, the gains for the proposed scheme are consistently better than the other schemes. Similarly, we plot the gains in normalized efficiency for a given normalized cell-edge throughput in Fig. 4.10. The proposed scheme achieves consistently higher gains in normalized efficiency than the other schemes, especially for high normalized cell-edge throughput.

### Efficiency-Jain Tradeoff Comparison

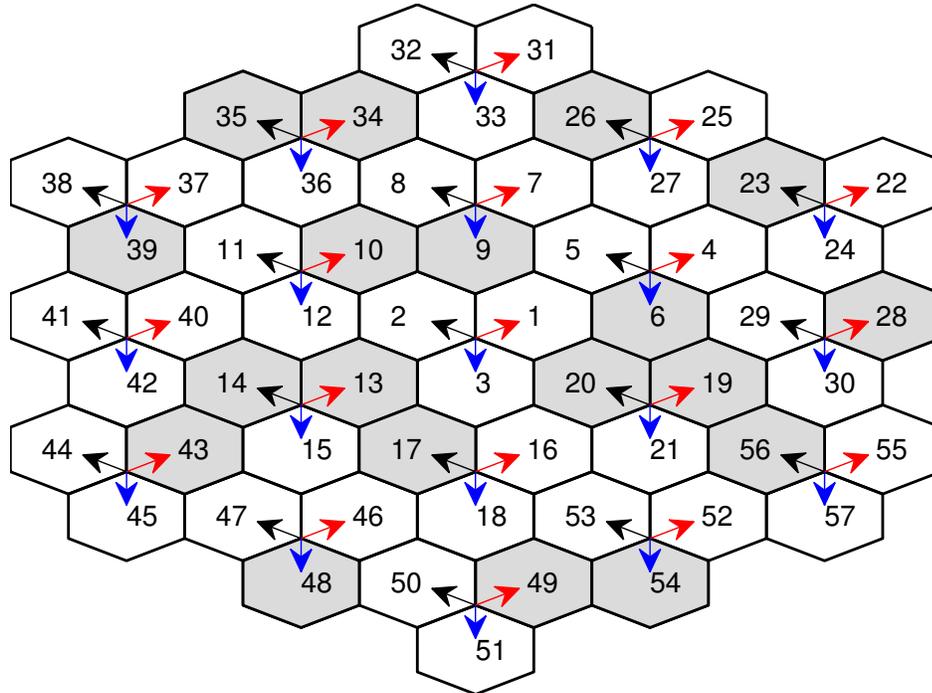
In this section, we compare the EJTs achieved by the proposed and the baseline schemes. To do so, we consider the case of uniform and nonuniform UT distribution.



**Figure 4.9:** Gain (%) in normalized cell-edge throughput versus normalized efficiency, as compared to reuse-1.



**Figure 4.10:** Gain (%) in normalized efficiency versus normalized cell-edge throughput, as compared to reuse-1.



**Figure 4.11:** Network layout with nonuniform user distributions, where each gray-sector has 10 UTs and each white-sector has 2 UTs.

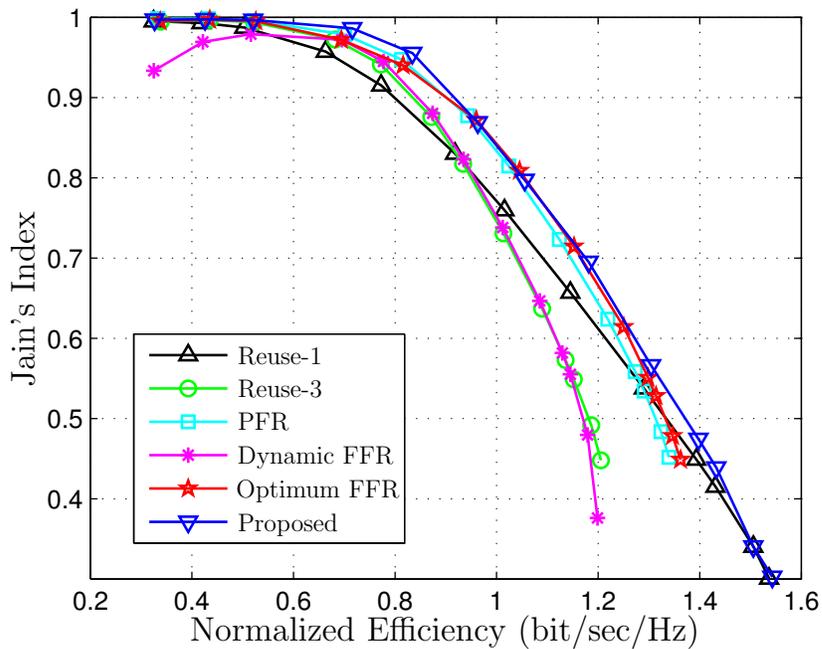
For the uniform UT distribution, each sector is assumed to have 10 UTs, as specified by IMT-Advanced guidelines and explained in Section 3.6.1. For the nonuniform UT distribution, we assume that one-third of the sectors are highly loaded with 10 UTs per sector, while two-thirds of the sectors are lightly loaded with 2 UTs per sector. The locations of the highly loaded and lightly loaded sectors are determined randomly and the resulted nonuniform layout is shown in Fig. 4.11, where the gray-sectors represent the highly loaded sectors and the white-sectors represent the lightly loaded sectors. Despite the simplicity of the considered nonuniform UT distribution model, it is sufficient to demonstrate the effectiveness of the proposed scheme in achieving network-wide fairness for nonuniform UT distributions, as we will see shortly.

In Figs 4.12(a) and 4.12(b), we plot the EJT<sub>s</sub> for the cases of uniform and nonuniform distributions, respectively, for the proposed and the baseline schemes. For all schemes, we use the optimal EJT scheduling policy that we proposed in Section 3.5.

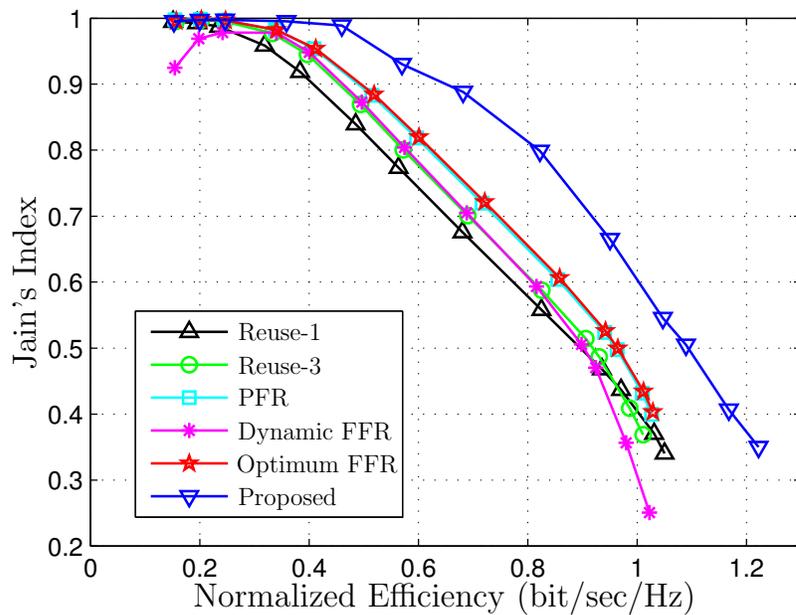
For the case of uniform UT distribution depicted in Fig. 4.12(a), it can be seen that reuse-3, PFR, dynamic FFR, optimum FFR, and the proposed scheme outperform reuse-1 in the high fairness region. However, reuse-3, PFR, dynamic FFR, and optimum FFR fail to provide an efficiency that is comparable to reuse-1 in the high efficiency region. Interestingly, the proposed scheme outperforms all the baseline schemes in both the high efficiency and high fairness regions. For the case of nonuniform UT distribution depicted in Fig. 4.12(b), it can be seen that the proposed scheme significantly outperforms all the baseline schemes in both the high efficiency and high fairness regions. This demonstrates that proposed scheme can effectively achieve network-wide fairness by coordinating resource allocations across multiple sectors.

To further examine the EJT achieved by the proposed and the baseline schemes, we plot in Figs 4.13(a) and 4.13(b) the gains in normalized efficiency achieved by the different schemes as compared to reuse-1, for the cases of uniform and nonuniform UT distributions, respectively. Similar to Figs 4.9 and 4.10, the gain in normalized efficiency for a particular scheme for a given Jain's index  $x$  is given by  $\text{Gain}_{\text{scheme}}(x) = \frac{\eta_{\text{scheme}}(x) - \eta_{\text{reuse1}}(x)}{\eta_{\text{reuse1}}(x)} \cdot 100\%$ , where  $\eta_{\text{scheme}}(x)$  is the normalized efficiency achieved by a particular scheme at Jain's index  $x$ , which can be obtained from the tradeoff curves given in Figs 4.12(a) and 4.12(b). It can be seen from these figures that the proposed scheme achieves consistently higher gains in normalized efficiency than the other schemes, especially for higher values of Jain's index. Moreover, the gains are more substantial for the case of nonuniform UT distribution depicted in Fig. 4.13(b); e.g., for a Jain's index of 0.9, the gain in normalized efficiency achieved by the proposed scheme is 60%.

Similar to Figs 4.13(a) and 4.13(b), we plot in Figs 4.14(a) and 4.14(b) the gains in Jain's index achieved by the different schemes as compared to reuse-1, for the cases of uniform and nonuniform UT distributions, respectively. For the case of uniform UT distributions, we can see that the gains in Jain's index are modest

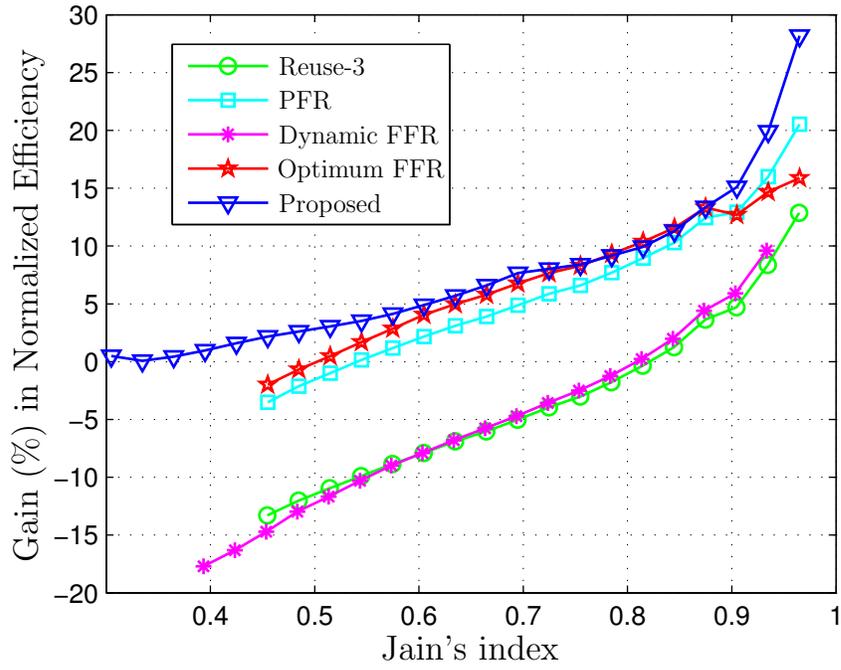


(a) Uniform UT distribution.

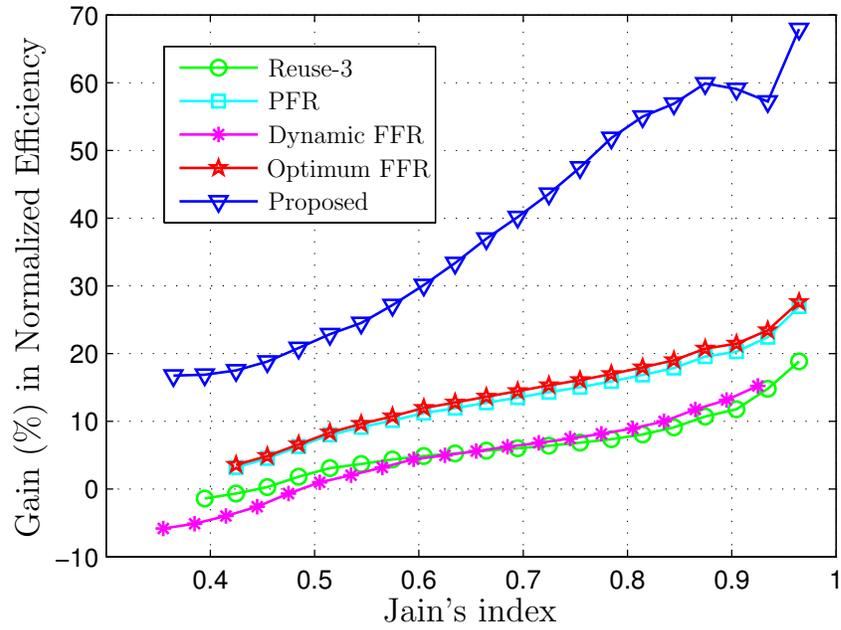


(b) Nonuniform UT distribution.

**Figure 4.12:** Jain's index versus normalized efficiency, for  $\beta \in \{0.03, 0.04, 0.05, 0.075, 0.1, 0.15, 0.2, 0.3, 0.5, 0.8, 1, 2, 5\}$  (bit/sec/Hz)



(a) Uniform UT distribution.



(b) Nonuniform UT distribution.

**Figure 4.13:** Gain (%) in normalized efficiency versus Jain's index, as compared to reuse-1.

for the proposed scheme, while the baseline schemes incur losses, especially for high normalized efficiency. On the other hand, for nonuniform UT distribution, the gains achieved by the proposed scheme are substantial; e.g., for a normalized efficiency of 0.9 bit/sec/Hz, the gain in Jain's index achieved by the proposed scheme is 43%.

### Outage Comparison

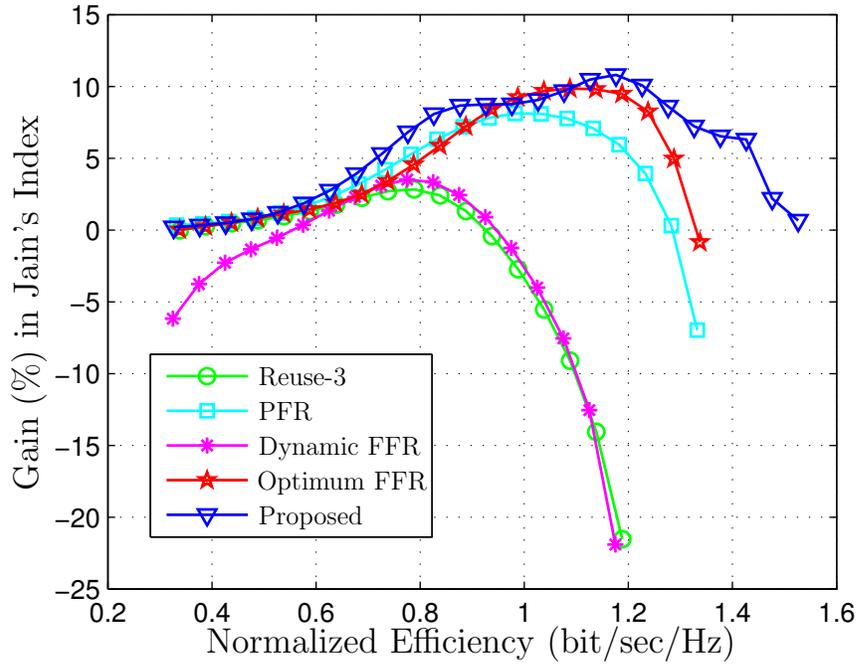
In Fig. 4.15, we plot the outage probability, which is defined as the probability of having the average UT throughput less than  $\bar{R}_{\min}$  for different schemes. It is clear from the figure that the proposed scheme achieves much lower outage probability as compared to other schemes. For example, at  $\bar{R}_{\min} = 0.05$  bits/sec/Hz, the proposed scheme has an outage probability that is at least 3.5 times less than those for PFR and optimum FFR, and at least 6 times less than those for reuse-1, reuse-3, and dynamic FFR.

### 4.6.3 Statistics of the Average Number of Restricted RBs

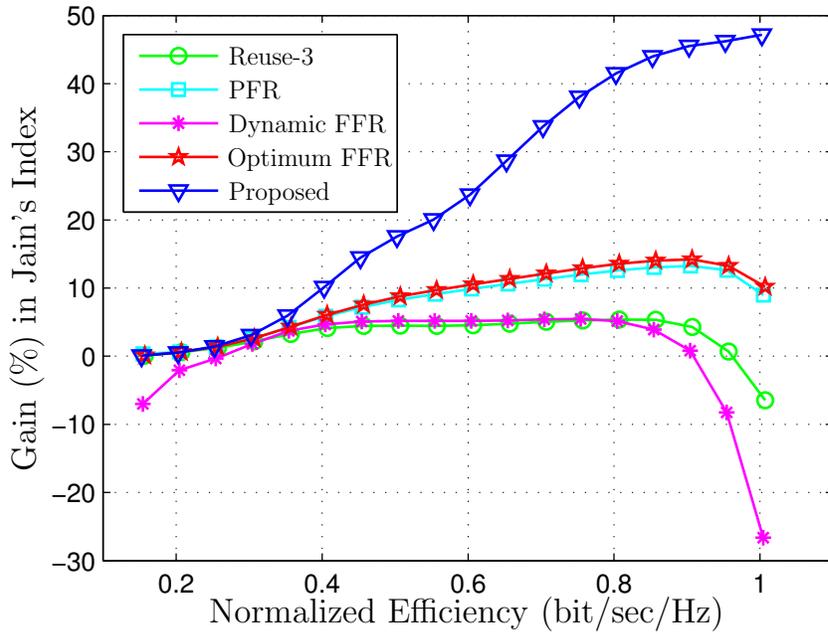
In Fig. 4.16, we plot the probability mass function of the average number of restricted RBs per sector for different fairness exponents. The proposed scheme has the flexibility to change the distribution of the restricted resources according to the desired fairness level. As  $\alpha$  increases, cell-edge users become more important and thus more resources need to be restricted, and vice versa. The figure shows also that the proposed scheme acts as reuse-1 (no restrictions) and reuse-3 (2/3 of RBs are restricted) with very small probabilities.

## 4.7 Conclusions

In this chapter, we proposed a novel ICIC scheme that runs in polynomial time and finds a near-optimum dynamic resource partitioning that maximizes a weighted

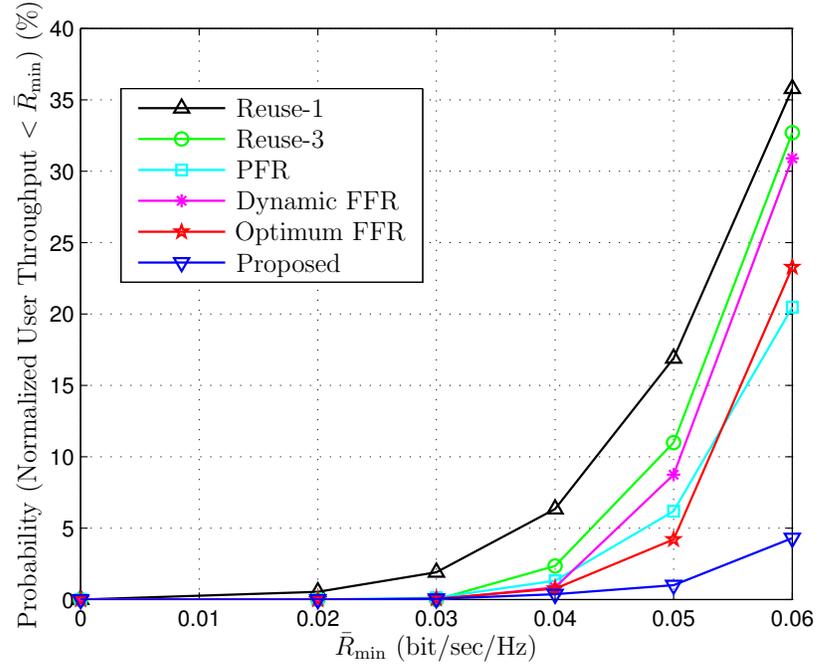


(a) Uniform UT distribution.

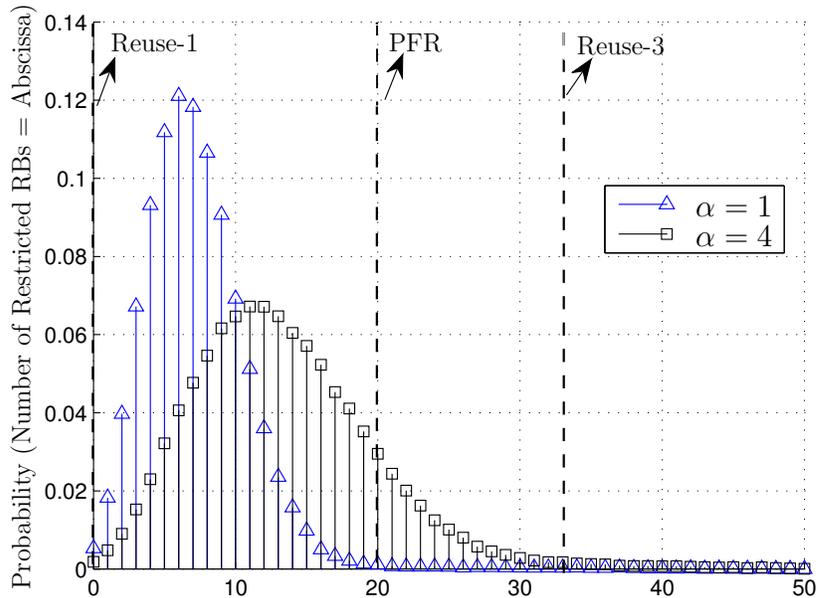


(b) Nonuniform UT distribution.

**Figure 4.14:** Gain (%) in Jain's index versus normalized efficiency, as compared to reuse-1.



**Figure 4.15:** Outage probability versus  $\bar{R}_{\min}$ , which is defined as the probability that the average UT throughput is less than  $\bar{R}_{\min}$ .



**Figure 4.16:** Probability mass function of the average number of restricted RBs per sector (the total number of RBs is 50).

sum-rate in the entire network in a distributed manner. The proposed algorithm can be used with any AMC scheme, including discrete-rate AMC schemes. The proposed scheme is developed using primal-decomposition method, which is utilized to decompose the problem into a master problem and multiple subproblems. The master problem is solved iteratively using projected-subgradient method. We reveal that each subproblem has a network flow structure which makes it amenable to powerful minimum cost network flow algorithms and thus results in significant reduction in the computational complexity. Simulation results show that the proposed scheme achieves high gain in sum-rate efficiency, Jain's fairness index, cell-edge throughput, and outage probability, as compared to reuse-1, reuse-3, partial frequency reuse, dynamic fractional frequency reuse, and optimum fractional frequency reuse.

## Chapter 5

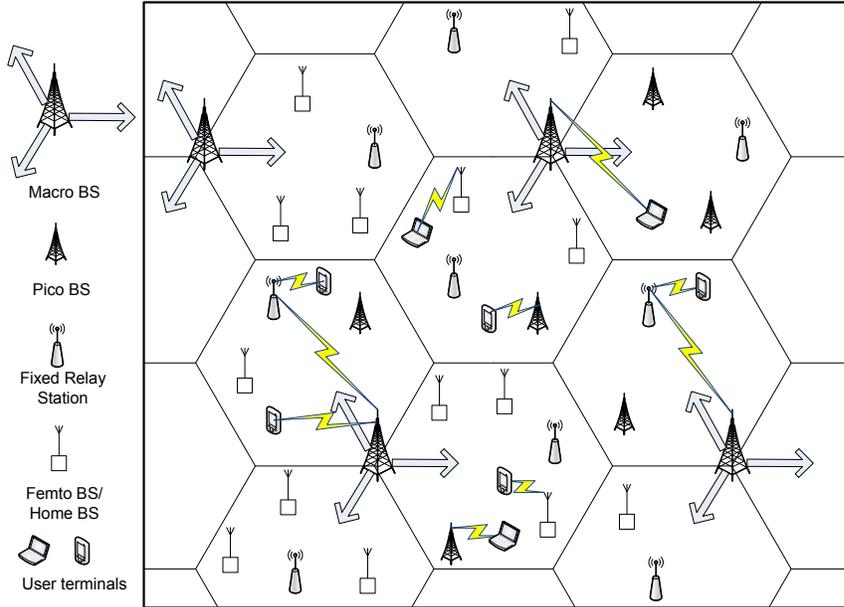
# Inter-cell Interference Coordination in Heterogeneous Networks

### 5.1 Introduction

As mentioned before, the demand for high-rate mobile data is expected to surge in the coming years (cf. [1]). Given the scarcity of radio resources, this surge in demand presents an unprecedented challenge to service providers. One approach to tackle this challenge is to aggressively reuse these scarce resources by increasing the density of conventional BSs, which are called macro-BSs herein. Unfortunately, such an approach is not economically feasible due to the high cost of macro-BSs. Moreover, increasing the density of macro-BSs in the same geographical region can lead to intolerable interference. In addition, acquiring new sites for mounting macro-BS's tower can be very challenging [93]. To overcome the drawback of high cost of dense deployment of macro-BSs, low-cost low-power BSs can be deployed in the same geographical coverage region of macro-BSs.

A network that consists of macro-BSs and a mixture of low-power BSs with different transmission powers is widely referred to as a heterogeneous network (Het-Net) [39, 93–98]. These low-power BSs can be classified into pico-BSs, femto-BSs,

and relay stations, as shown in Fig. 5.1. A brief description of these low-power BSs is given below.



**Figure 5.1:** Illustration of a HetNet.

**Pico-BS:** Pico-BSs are deployed by the service provider in more or less ad-hoc manner to cover coverage holes and/or to satisfy high demand in hot-spots [39]. It is designed to be deployed both outdoor and indoor. An LTE pico-BS has a maximum transmit power of 24 to 37 dBm [95, 99].

**Femto-BS:** Femto-BS (or Home BS) is a low-cost, low-power, plug-and-play BS that can be installed in an indoor environments and it is usually purchased and maintained by a subscriber. Due to the short distance, a femto-BS can provide high throughput with low power, through the reuse of the licensed spectrum and it is an excellent solution for indoor coverage problem, e.g. coverage inside buildings. These femto-BSs are connected to the network through a broadband connection such as DSL. Femto-BSs are expected to have a maximum transmit power of 20 to 24 dBm [95, 99]. As explained in [100], there are three access methods for femto-BSs:

- Open Access: all users have full access to the femto-BS.
- Closed Access: only subscribed users can access the femto-BS.
- Hybrid Access: subscribed users have full access to the femto-BS while non-subscribed users have access to only limited resources offered by the femto-BS.

Open access strategy is the most efficient strategy from capacity and coverage point of views, as it allows macro-cell users to use the femto-BS in coverage holes. However, a major drawback of this strategy is that the owners of the femto-BSs will not have incentives to let other non-subscribed users to use their femto-BSs without sharing the revenue. Consequently, closed access will be most likely the most widely deployed strategy.

**Relay station:** Unlike pico-BSs and femto-BSs, relay stations have wireless backhaul connection that is provided by the macro-BS. In LTE, a relay station has a maximum transmit power of 30 to 37 dBm [99].

HetNet is a paradigm shift in the way we perceive, design, and optimize cellular networks. Seven fundamental differences between conventional (homogeneous) networks and HetNets are highlighted in [93]. In the following, we highlight the two most important differences that are relevant to our work [93]:

- Topology: in analyzing and simulating homogeneous networks, conventional macro-BSs are assumed to be placed in a hexagonal-grid. While this assumption also holds for macro-BSs in HetNets, it does not hold for low-power BSs such as pico-BSs, femto-BSs, and relay stations. This is because these low-power BSs are deployed in almost ad-hoc manner without prior planning, in order to eliminate coverage holes or to satisfy surge in demands in hot-zones [39, 96]. This results in irregular cell shapes, which raises challenges for semi-static interference management schemes that depend on regular cell shapes such as FFR schemes.

- Cell association: conventionally, UTs are associated to the BS to which it has the highest wideband SINR (cf. Section 3.6.2). However, since the transmit power of low-power BSs is much lower than that of a macro-BS, the coverage region of these low-power BSs can be very small in comparison to that of macro-BSs. This results in associating most of the users to macro-BSs which defies the purpose of low-power BSs. In order to balance the load across the BSs, range extension is applied to low-power BSs through biasing. In particular, in the process of associating UTs to BSs, a bias value (virtual) is added to the SINR received from low-power BSs in order to make low-power BSs more preferable to UTs [93, 97]. Range extension is effective since most low-power BSs are lightly loaded; however, this can potentially results in very low SINRs for UTs who are associated to low-power BSs. This raises the need for new interference management schemes that can handel such high interference environment; this is addressed in the following section.

To highlight the interference management aspect of HetNet, we consider only HetNets that consist of macro-BSs and pico-BSs.

## 5.2 ICIC schemes in Heterogeneous Networks

In this section, we explain the baseline ICIC schemes that can be used in the context of HetNets. We also discuss the application of the scheme we proposed in Chapter 4 and its variants to HetNets.

### 5.2.1 Baseline schemes

The baseline schemes we explained in Section 4.5 can also be adopted in HetNets. Reuse-1, in which all resources are used by all BSs, can be applied in straight-forward manner for both macro-BSs and pico-BSs. On the other hand, reuse-3, PFR [42],

dynamic FFR [46], and optimum FFR [55] can be applied only to macro-BSs while reuse-1 is used for pico-BSs<sup>1</sup>. However, as alluded in [93,101], these schemes may not be very effective in HetNets as we will also verify hereafter.

A scheme that is more powerful in mitigating interference in HetNets is to implement ICIC in time-domain, which is referred to as enhanced ICIC (eICIC) in 3GPP [93,101]. An advantage of eICIC is that it provides interference-protection for both data and control channels [97,98,101,102]. This is implemented by periodically restricting macro-BSs from using particular sub-frames in order to reduce the interference experienced by UTs that are associated with pico-BSs. These restricted sub-frames are called almost-blank-sub-frames (ABS), where the word “almost” is used since macro-BS may still transmit control messages in these sub-frames [39]. For example, 1/4 ABS scheme means that one-sub-frame out of every four sub-frames is blanked and reserved for pico-BSs.

### 5.2.2 Proposed schemes

In Chapter 4, we proposed a distributed ICIC scheme in the context of homogeneous network deployment, where the access network consists solely of macro BSs that are equipped with a tri-sector antenna. In this section, we discuss the application of the proposed ICIC scheme in HetNets. Since the proposed scheme is applicable to any cell shapes and any type of access points, the proposed scheme can be used in HetNets in straightforward manner, as long as communication is possible between the different access points types, which can be realized in practice. For example, in 3GPP Release 10 (LTE-Advanced) [103], macro-pico BS communication is possible through the X2 interface and macro-femto BS communication is possible through the

---

<sup>1</sup>It was observed through extensive simulations that applying reuse-3, PFR, dynamic FFR, and optimum FFR for both macro-BSs and pico-BSs can degrade the performance significantly. Such a degradation is a consequence of irregular shapes of pico-cells and the high disparity between the loads on macro-BSs and pico-BSs. As a result, these schemes are applied only to macro-BSs herein.

OAM (Operations, Administration, and Maintenance) interface [62]. The proposed algorithm can be executed every  $\tau$  sub-frames, where  $\tau$  can be adjusted to suit the practical limitations of X2 and OAM interfaces. However, there are two important implications of applying the proposed ICIC scheme in HetNet:

1. Bound optimization: in homogeneous networks, it was argued and verified by simulating a small network that bound optimization is a good approach in approaching near-optimality. This was mainly due to the fact that UTs (especially cell-edge UTs) are mainly affected by the most dominant restricted interferers. However, due to the dense deployment of BSs in HetNets, UTs are potentially affected by more than one BS, which makes the bound optimization not necessarily near-optimum.
2. Due to the ad-hoc nature of deploying pico-BSs, fast communication links between macro-BSs and pico-BSs may not be readily available which raises the need for reducing the message exchange between BSs.

The first implication can be addressed by executing the algorithm more than once. That is, after finding the optimum  $\{I_n^{(k)}\}$  in one run, the algorithm sets  $H_{m,n}^{(\tilde{k},k)} = 0$ ,  $\forall m, n, k, \tilde{k} | I_n^{(k)} = 1$ , and executes the algorithm again to handle the second most dominate restricted interferer. For homogeneous networks, we observed that executing the algorithm more than once provides only an incremental gain, cf. Table 4.1. As we will see later, executing the algorithm more than once provides substantial gains in HetNets.

To address the second implication, we devise a variant of the proposed scheme that is inspired by the ABS concept; we call this scheme the proposed dynamic ABS.

### 5.2.3 Proposed Dynamic ABS Scheme

We changed the formulation of the ICIC scheme presented in Chapter 4 such that it resembles an ABS scheme. This is achieved by assuming that the  $I$  value is the same for all RBs in a particular sub-frame. In other words, in every sub-frame, either all RBs are transmitted with full power or all RBs are muted. The optimization problem can be formulated as follows:

$$\begin{aligned} & \underset{\{x_{m,n}^{(k)}\}, \{I^{(k)}\}}{\text{maximize}} && \sum_{k=1}^K \sum_{m=1}^M \alpha_m^{(k)} x_{m,n}^{(k)} R_{m,n}^{(k)} \end{aligned} \quad (5.1a)$$

$$\text{subject to} \quad \sum_{m=1}^M x_{m,n}^{(k)} = 1 - I^{(k)}, \quad \forall k, \quad (5.1b)$$

$$x_{m,n}^{(k)}, I^{(k)} \in \{0, 1\}, \quad \forall m, k, \quad (5.1c)$$

where the binary variables  $I^{(k)}$  are introduced such that  $I^{(k)} = 1$  when the sub-frame is muted in sector  $k$  and  $I^{(k)} = 0$  otherwise.

The only difference between this formulation and the one we considered in Chapter 4 is that there is only one  $I^{(k)}$  variable for each sector for each sub-frame (recall that in Chapter 4 we had  $I_n^{(k)}$  for each RB, in each sector, in each sub-frame). With this formulation, we can use the same derivation steps we used before, and the algorithm stays the same with one important exception: Each BS exchanges the sum of the lagrange multipliers over all RBs instead of exchanging all lagrange multipliers for all RBs. This significantly reduces the information exchange between BSs by a factor of  $N$ . However, dynamic ABS is expected to perform worse than muting RB-by-RB as we loose the granularity and frequency diversity. This will be verified in the simulation results in Section 5.4.

The proposed dynamic ABS scheme can be distinguished from the conventional ABS scheme in the following ways. First of all, in the proposed scheme not all

BSs have to be muted at the same sub-frame. Moreover, muting sub-frames in the proposed scheme is not periodic and it changes as a function of the load and the desired efficiency-fairness tradeoff. Finally, in the proposed scheme pico-BSs can also be muted.

### 5.3 Simulation Setup

We consider a heterogeneous cellular network based on the 3GPP guidelines for case 6.2 scenario [99]. As per these guidelines, a hexagonal layout with wraparound is considered with 57 hexagonal macro sectors and 25 UTs per macro sector. These macro sectors are served by 19 macro-BSs, each with a tri-sector antenna to serve a 3-sector cell-site. In each macro sector,  $N_{\text{pico}}$  pico-BSs are randomly dropped, while making sure that the minimum distance between pico-BSs and macro-BSs satisfy the prescribed limits in [99]. The users are uniformly dropped in the 57 sectors, and Monte Carlo simulations are carried over  $10^3$  sub-frames and averaged over 5 independent drops.

The users are associated either to a macro-sector antenna or to a pico-BS based on the average received power, taking into account a range extension margin of 6 dB for pico-BSs as explained before. The average received power depends on large channel variations, which are mainly due to non-uniform antenna patterns and distance-dependent path-loss and correlated shadowing. To incorporate these variations in our simulations, we considered both LOS and NLOS users as in [99]. For the LOS macro-UTs, the path-loss exponent is set to be 2.42 and the shadowing is assumed log-normal with standard deviation of 4 dB. For the LOS pico-UTs, the path-loss exponent is set to be 2.09 and the shadowing is assumed log-normal with standard deviation of 10 dB. For the NLOS macro-UTs, the path-loss exponent is set to be 4.28 and the log-normal shadowing standard deviation is set to be 6 dB. For the NLOS

pico-UTs, the path-loss exponent is set to be 3.75 and the log-normal shadowing standard deviation is set to be 10 dB. The system parameters are based on the LTE standard, whereby each RB is composed of a time slot of 7 OFDM symbols and 12 subcarriers [29]. The simulation parameters are given in Tables 5.1 and 5.2.

In addition to large variations, the received signal power is subject to small-scale variations due to time-varying multipath fading. Variations of the received signal power due to fading within each RB is negligible and hence, the channels can be assumed fixed over each RB. For other RBs the channels take on different values depending on the spectro-temporal correlation of the IMT-Advanced model for the UMa scenario [35].

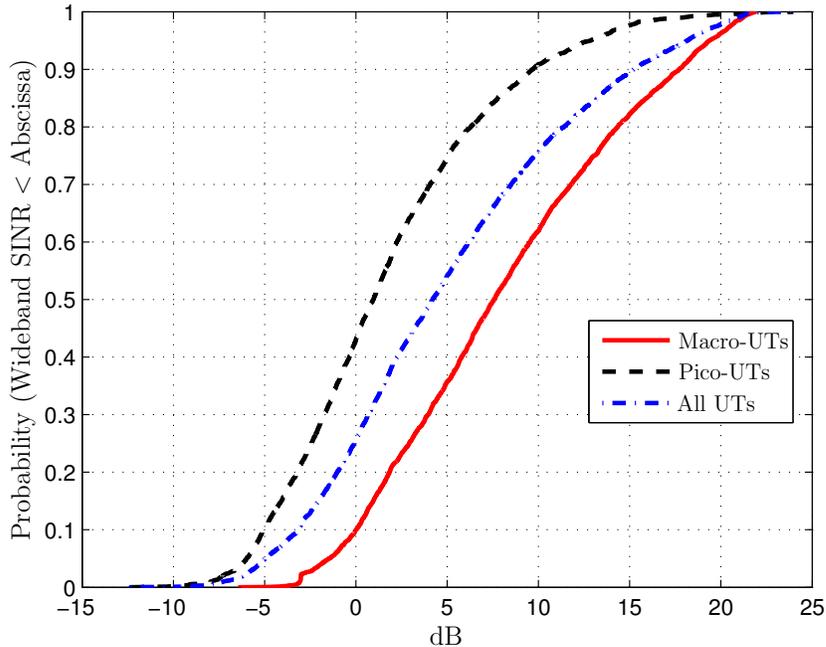
At sub-frame  $t$ , each user calculates its SINR on each RB, which involves the calculation of the interference from all macro-BSs and pico-BSs that are using the same RB. Based on the calculated SINRs, each user determines the data rate that can be reliably communicated on every RB using the AMC Table 3.2. The set of all the rates at sub-frame  $t$  are transmitted through feedback control channels to the serving BS, which subsequently determines the appropriate scheduling variables, depending on previously scheduled rates and the adopted scheduling strategy.

### 5.3.1 Wideband SINR CDF

Using the 3GPP case 6.2 model with the above parameter settings, we plot the CDF of the wideband SINR experienced by macro-UTs and pico-UTs in Fig. 5.2 for 4 pico-BSs per macro-sector. From this figure, it is clear that the SINR for pico-UTs is much less than their macro counterparts, mainly due to range extension and the low transmitting power of pico-BSs.

**Table 5.1:** Simulation parameters based on 3GPP heterogeneous network deployment Scenario 6.2.

Parameter	Assumption
Number of macro sectors	57 (wraparound)
Number of UTs per sector	25
Inter-site distance	500 m
Macro BS's height	32 m
Min. distance between a UT and a macro-BS	35 m
Min. distance between a UT and a pico-BS	10 m
Min. distance between a pico-BS and a macro-BS	75 m
Min. distance between any two pico-BSs	40 m
UT's speed	3 km/h
Bandwidth (downlink)	10 MHz
Sub-carrier spacing	15 KHz
Number of RBs ( $N$ )	50
OFDM symbol duration	66.67 $\mu$ s
Number of sub-carriers per RB	12
Number of OFDM symbols per RB	7
Number of drops	5
Number of sub-frames per drop	1000
Noise power per RB ( $P_n$ )	-114.45 dBm
Carrier frequency	2.0 GHz
Total macro-BS's transmit power	46 dBm
Total pico-BS's transmit power	24 dBm
Path loss for LOS macro-UTs (dB)	$103.4 + 24.2 \log_{10}(d(\text{km}))$
Path loss for NLOS macro-UTs (dB)	$131.1 + 42.8 \log_{10}(d(\text{km}))$
Path loss for LOS pico-UTs (dB)	$103.8 + 20.9 \log_{10}(d(\text{km}))$
Path loss for NLOS pico-UTs (dB)	$145.4 + 37.5 \log_{10}(d(\text{km}))$
Penetration loss	20 (dB)
Shadowing standard deviation for LOS macro-UTs	4 (dB)
Shadowing standard deviation for NLOS macro-UTs	6 (dB)
Shadowing standard deviation for LOS & NLOS pico-UTs	10 (dB)
Shadowing correlation distance	50 m
Smoothing factor ( $\mu$ )	0.01



**Figure 5.2:** CDF of the wideband SINR of macro-UTs and pico-UTs.

## 5.4 Simulation Results

We now present the performance results using the 3GPP case 6.2 model with the above parameter settings assuming a proportional fair scheduler (i.e.,  $\alpha$ -fair scheduler with  $\alpha = 1$ ). In this section, efficiency of macro-UTs (pico-UTs) is defined as the sum of the long-term average rates of all macro-UTs (pico-UTs) in the network. Similarly, efficiency of all UTs is defined as the sum of the long-term average rates of all UTs (i.e., both macro-UTs and pico-UTs) in the network.

### 5.4.1 Fixed Number of Pico-BSs and Fixed Range Extension Margin

In this section, we compare the performance of the proposed schemes with the baseline schemes presented in Section 5.2.1, assuming 4 pico-BSs per macro sector and

**Table 5.2:** Simulation parameters based on 3GPP heterogeneous network deployment Scenario 6.2 (cont'd).

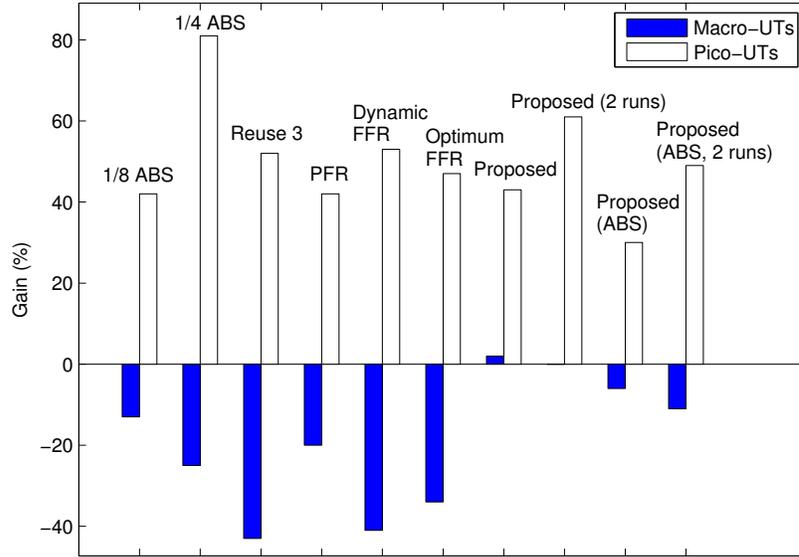
Parameter	Assumption
Channel estimation delay	4 sub-frames
SINR estimation margin	6 dB
Small-scale fading model	IMT-Advanced channel model [35]
Traffic model	Full buffer
Macro-BS's antenna gain (boresight)	17 dBi
Pico-BS's antenna gain (boresight)	5 dBi
User antenna gain	0 dBi
Macro-BS's antenna tilt ( $\phi_t$ )	15° [99]
Macro-BS's horizontal antenna pattern	$A(\theta) = -\min [12(\frac{\theta}{70^\circ})^2, 25 \text{ dB}]$ [99]
Macro-BS's elevation antenna pattern	$A_e(\phi) = -\min [12(\frac{\phi-\phi_t}{10^\circ})^2, 20 \text{ dB}]$ [99]
Macro-BS's combined antenna pattern	$-\min [-(A(\theta) + A_e(\phi)), 25 \text{ dB}]$ [99]
Pico-BS's combined antenna pattern	$A(\theta, \phi) = 0 \text{ dB}$

a range extension margin of 6 dB for pico-BSs. Based on these simulation parameters, 54.7% of the UTs are associated to macro-BSs while 45.3% of the UTs are associated to pico-BSs, on average.

### Efficiency Comparison

In Fig. 5.3, we plot the gains in efficiency of macro-UTs and pico-UTs, as compared to reuse-1, for the proposed schemes and the baseline schemes. It can be seen from this figure that, while the baseline schemes achieve high gains for pico-UTs, they incur high losses for macro-UTs. For example, 1/4 ABS achieves a gain of 81% for pico-UTs at the expense of a loss of 25% for macro-UTs. On the other hand, the proposed schemes achieve high gains for pico-UTs at the expense of small or no loss for macro-UTs. We can also observe from this figure that running the proposed algorithm twice yield significant gains compared to running it once, unlike the case of homogeneous networks that we addressed in Chapter 4. In addition, we note that the ABS versions

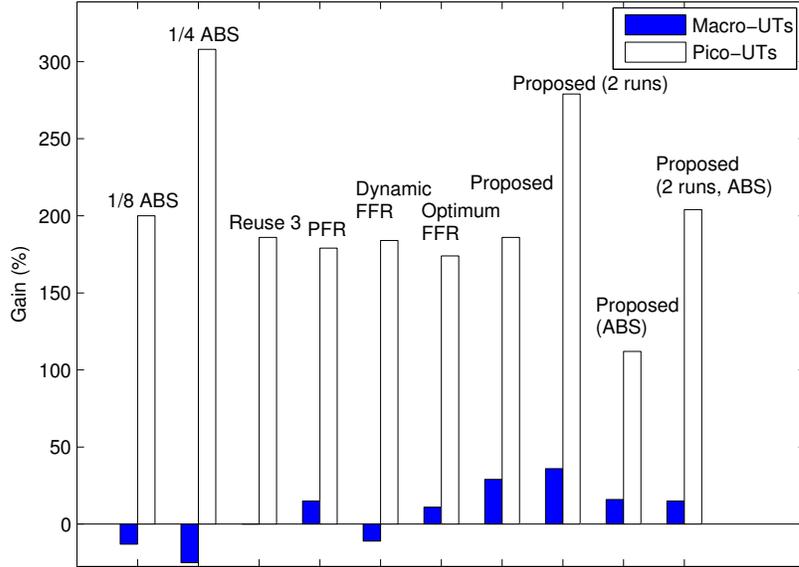
of the proposed scheme performs worse than the proposed scheme in order to reduce the complexity of message exchange.



**Figure 5.3:** Gain (%) in efficiency of macro-UTs and pico-UTs, as compared to reuse-1.

### Cell-Edge Throughput Comparison

Similar to Chapter 4, the cell-edge throughput is defined as the 5<sup>th</sup> percentile of the user throughput. In Fig. 5.4, we plot the gains in the cell-edge throughput for the proposed schemes and the baseline schemes as compared to reuse-1, for macro-UTs and pico-UTs. While 1/4 ABS achieves the highest gain for pico-UTs (%308), this large gain comes at the expense of a loss of 25% for macro-UTs. On the other hand, the proposed scheme (2 runs) achieve gains of 279% for pico-UTs and 36% for macro-UTs.



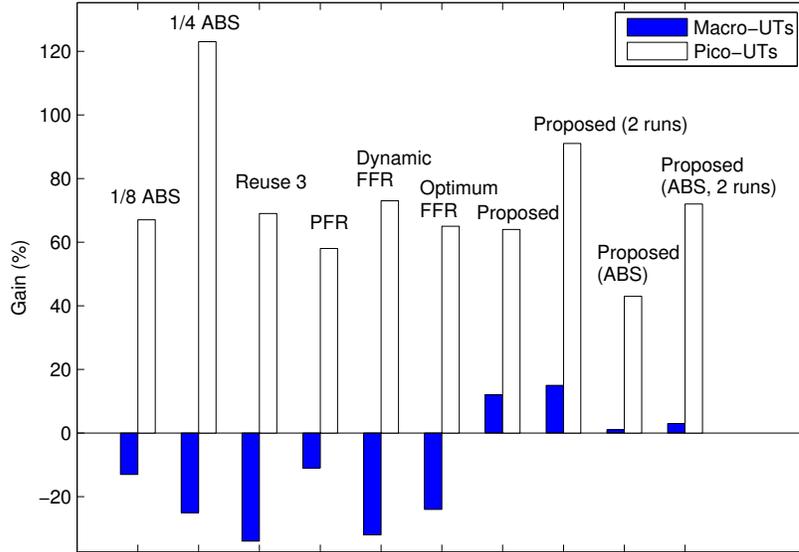
**Figure 5.4:** Gain (%) in cell-edge throughput for macro-UTs and pico-UTs, as compared to reuse-1.

### Median Throughput Comparison

In Fig. 5.5, we plot the gains in the median throughput for the proposed schemes and the baseline schemes as compared to reuse-1, for macro-UTs and pico-UTs. It can be seen from this figure that, while all schemes achieve high gains for pico-UTs, the baseline schemes incur losses in the median throughput for macro-UTs. In particular, 1/8 ABS, 1/4 ABS, reuse-3, PFR, dynamic FFR, and optimum FFR incurs losses of 13%, 25%, 34%, 11%, 32%, and 24% for macro-UTs, respectively. On the contrary, proposed, proposed (2 runs), proposed (ABS), and proposed (2 runs, ABS) achieve gains of 12%, 15%, 1%, and 3% for macro-UTs.

### Fairness Comparison

In Fig. 5.6, we plot the gains in Jain's fairness index for the proposed schemes and the baseline schemes as compared to reuse-1, for macro-UTs and pico-UTs. It

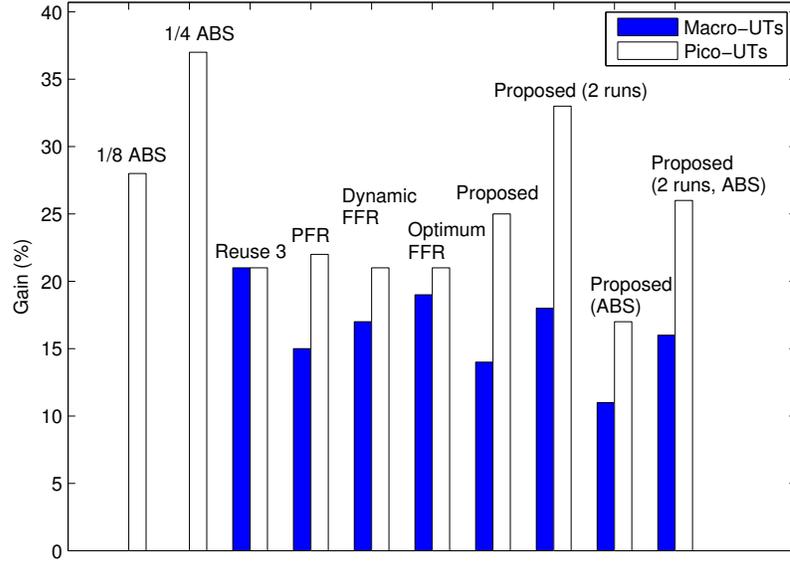


**Figure 5.5:** Gain (%) in median throughput for macro-UTs and pico-UTs, as compared to reuse-1.

can be seen from this figure that all schemes improve fairness as measured by Jain’s index for pico-UTs. Moreover, except for 1/8 ABS and 1/4 ABS, all other schemes improve fairness as measured by Jain’s index for macro-UTs.

### 5.4.2 Variable Number of Pico-BSs and Fixed Range Extension Margin

In this section, we examine the performance of the proposed and the baseline schemes as a function of the number of pico-BSs. To do so, we vary the number of pico-BSs per sector from 1 to 4, and for each case, we examine the gains achieved by the proposed and the baseline schemes in efficiency, cell-edge throughput, median throughput, and Jain’s fairness index. Unlike the previous section where we examined macro-UTs and pico-UTs separately, in this section we examine the performance of all UTs combined since the percentage of pico-UTs vary based on the number of



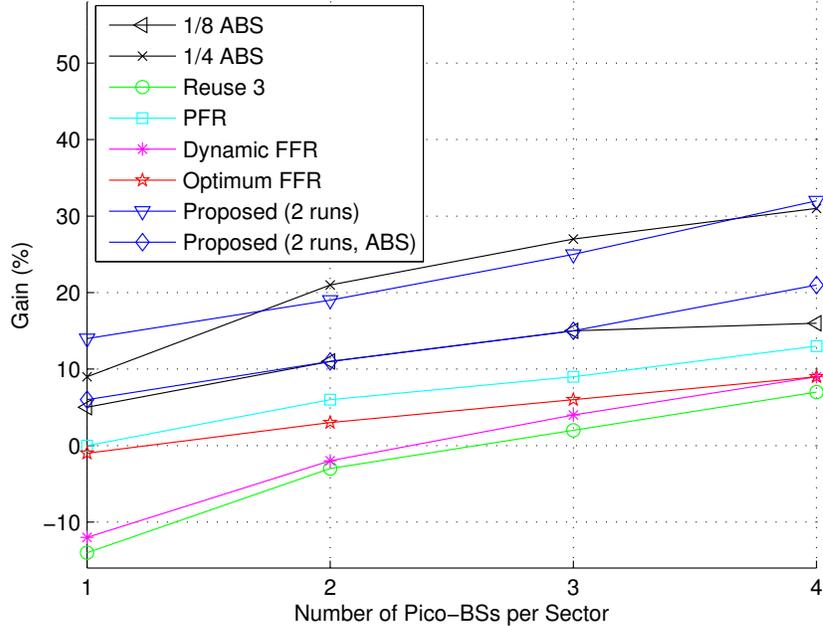
**Figure 5.6:** Gain (%) in Jain's index for macro-UTs and pico-UTs, as compared to reuse-1.

pico-BSs per sector. In particular, based on the considered simulation parameters, the percentages of pico-UTs are 16.4%, 27.2%, 37.5%, and 45.3% for 1, 2, 3, and 4 pico-UTs per sector, respectively. Since we found in the previous section that the proposed schemes perform best if they are executed twice, we only include the results of proposed (2 runs) and proposed (2 runs, ABS). Finally, we assume a fixed range extension margin of 6 dB for pico-BSs.

### Efficiency Comparison

In Fig. 5.7, we plot the gains in efficiency achieved by the baseline and the proposed schemes as compared to reuse-1, for different number of pico-BSs per sector. The general trend for all schemes is that as the number of pico-BSs per sector increases, interference increases, which leads to an increase in the gain achieved by interference management schemes. The highest gains are achieved by 1/4 ABS and proposed (2

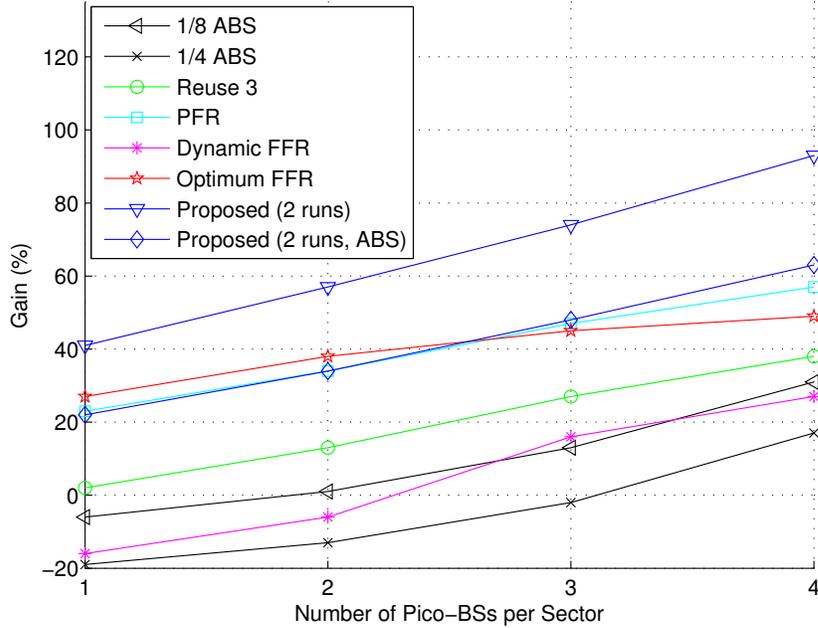
runs) schemes. However, as we will see shortly, the gains achieved by 1/4 ABS comes at significant degradation in the cell-edge throughput. We also note that Reuse 3 and dynamic FFR incur losses in the efficiency when the number of pico-BSs is 1 and 2.



**Figure 5.7:** Gain (%) in efficiency of all UTs as compared to reuse-1 for different number of pico-BSs.

### Cell-Edge Throughput Comparison

In Fig. 5.8, we plot the gains in the cell-edge throughput for the proposed schemes and the baseline schemes as compared to reuse-1. It can be seen from this figure that 1/4 ABS has the worst cell-edge performance. This is a consequence of reducing the resources for macro-UTs by 25% which reduces the throughput of cell-edge macro-UTs. Interestingly, proposed (2 runs) provides the highest gain in cell-edge throughput while maintaining a high gain in efficiency (cf. Fig. 5.7). Proposed (2 runs, ABS) provides a good compromise between complexity and gains in efficiency and cell-edge throughput.



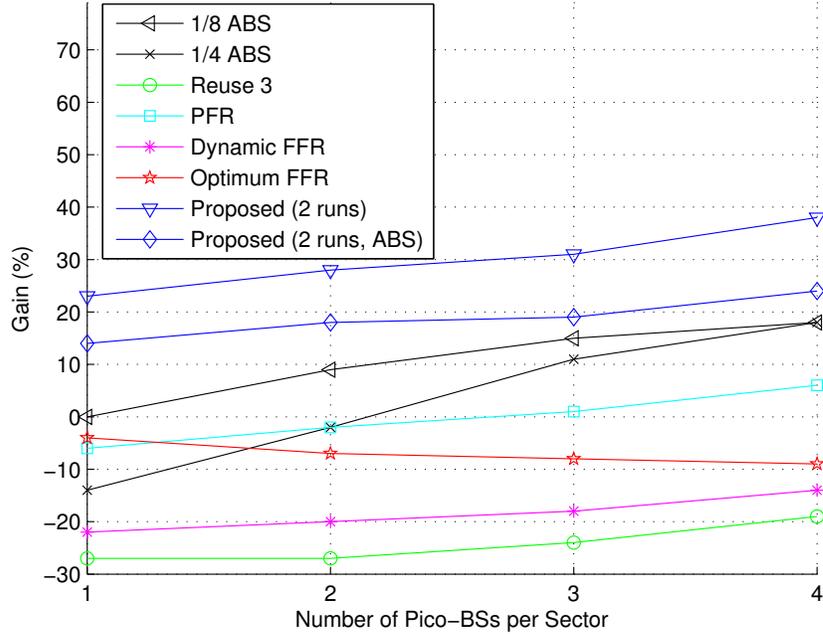
**Figure 5.8:** Gain (%) in cell-edge throughput for all UTs as compared to reuse-1 for different number of pico-BSs.

### Median Throughput Comparison

In Fig. 5.9, we plot the gains in the median throughput for the proposed schemes and the baseline schemes as compared to reuse-1. For the case of 1 pico-BS per sector, it can be seen from this figure that all the baseline schemes either provide no gain or incur losses in the median throughput as compared to reuse-1. Again, the proposed schemes perform very well in the median throughput.

### Fairness Comparison

In Fig. 5.10, we plot the gains in Jain's fairness index for the proposed schemes and the baseline schemes as compared to reuse-1. For the cases of 1 and 2 pico-BSs per sector, it can be seen from this figure that all the baseline schemes incur significant losses in fairness as compared to reuse-1. This comes as a consequence of reducing the resources available for macro-UTs. On the other hand, proposed (2 runs) provides

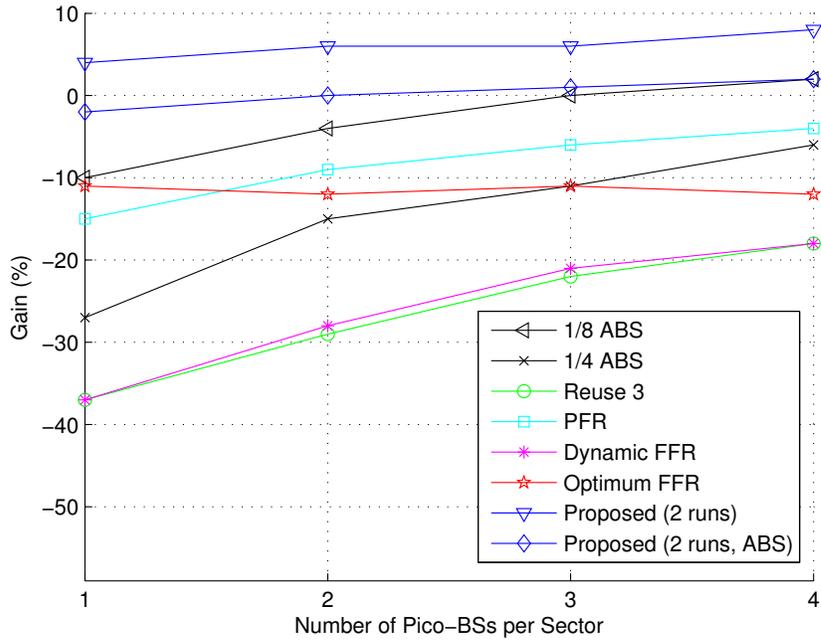


**Figure 5.9:** Gain (%) in median throughput for all UTs as compared to reuse-1 for different number of pico-BSs.

modest gains in fairness while proposed (2 runs, ABS) incur a small loss for the case of 1 pico-BS per sector.

## 5.5 Conclusions

Inspired by the ICIC scheme we proposed in Chapter 4 for homogeneous networks, we proposed variants of it in this chapter that are more suitable for HetNets. In one variant, we proposed to execute the ICIC algorithm more than once, since in HetNets, UTs are potentially affected by more than one BS, due to the dense deployment of BSs. This allows us to handle more than one dominant restricted interferer. In another variant, we proposed a dynamic ABS scheme in order to reduce the message exchange between BSs. Unlike most of the existing schemes, we demonstrated that it is possible to improve the performance of macro-UTs and pico-UTs, simultaneously, at



**Figure 5.10:** Gain (%) in Jain's fairness index for all UTs as compared to reuse-1 for different number of pico-BSs.

the expense of fast message exchange, by executing the proposed ICIC algorithm more than once. This includes improvement in sum-rate efficiency, cell-edge throughput, median throughput, and Jain's fairness index. We also showed that the proposed dynamic ABS scheme is a low-complexity scheme that provides a good compromise between complexity and gains in performance.

## Chapter 6

# Conclusions and Future Work

## 6.1 Summary and Contributions

In this thesis, we investigated efficiency-fairness tradeoff in interference-limited cellular networks. The contributions of this thesis can be summarized as follows:

- We showed that, in general, the commonly-used  $\alpha$ -fair policy [3] does not yield the optimal EJT's except for the two-user case. In particular, when the number of users is greater than two, it is shown that the gap between the efficiency achieved by the  $\alpha$ -fair policy and that achieved by the optimal EJT policy for the same Jain's index can be unbounded.
- We developed two procedures to achieve the optimal EJT's in the general case. In the first procedure, the set of feasible benefits is arbitrary, but finding the allocations that achieve the optimal EJT's involves solving potentially difficult optimization problems. In contrast, in the second procedure, the set of feasible benefits is assumed to have a monotonic tradeoff property that arises in many practical scenarios. This property is exploited to facilitate the search for allocations that achieve the optimal EJT's. We also provided sufficient conditions which, when satisfied, ensure that a given set  $\mathcal{C}$  possesses this property. These

sufficient conditions are given below:

- i.  $\mathcal{C}$  is convex;
- ii.  $x_{\min} \mathbf{1}_M \in \mathcal{C}$ ; and
- iii. every  $\mathbf{x} \in \mathcal{C}$  satisfies  $\mathbf{x} \succeq x_{\min} \mathbf{1}_M$ ,

where  $x_{\min} \geq 0$  provides a guarantee on the minimum benefit that each user receives.

- Using the sufficient conditions, we demonstrated that the monotonic tradeoff property arises naturally in various resource allocation problems, including the allocation of radio resources in the downlink of cellular networks. By exploiting this property, we devised computationally-efficient schedulers that achieve the optimal EJT in quasi-static and ergodic time-varying communication scenarios.
- Through analysis and extensive simulations, we compared the performance of the proposed optimal EJT scheduler,  $\alpha$ -fair scheduler, and MTMR scheduler. We found that the proposed optimal EJT scheduler outperforms both  $\alpha$ -fair and MTMR schedulers, in EJT, Efficiency-Median Tradeoff, and user satisfaction. On the other hand, the MTMR scheduler outperforms both the proposed optimal EJT and the  $\alpha$ -fair schedulers in Efficiency-Cell-edge Tradeoff; however, the MTMR scheduler incurs significant loss in EJT, Efficiency-Median Tradeoff, and user satisfaction.
- We proved that the only fairness function that is homogeneous of degree zero and concave is the constant function. In other words, it is not possible to construct a meaningful fairness function that is also concave.
- When the sufficient conditions are satisfied, optimal EJT benefit vectors were shown to have an interesting geometric interpretation. An optimal EJT benefit

vector,  $\mathbf{x}_\sigma^*$ , is the unique Euclidean projection of the equal allocation vector  $\frac{\sigma}{M}\mathbf{1}_M$  onto the set  $\{\mathbf{x}|\eta(\mathbf{x}) = \sigma, \mathbf{x} \in \mathcal{C}\}$ . In other words, a benefit vector  $\mathbf{x}_\sigma^*$  achieves the optimal EJT if there is no feasible benefit vector  $\mathbf{y} \neq \mathbf{x}_\sigma^*$  such that  $\eta(\mathbf{y}) = \sigma$  is closer to the fairest solution  $\frac{\sigma}{M}\mathbf{1}_M$ .

- We showed that in some scenarios, where the sufficient conditions are not satisfied, the tradeoff between efficiency and Jain's fairness index is not meaningful, i.e., it is possible to increase both efficiency and Jain's index at the same time.
- We demonstrated that the derived sufficient conditions are satisfied in the problem of congestion control in elastic traffic communication networks. Hence, the optimal EJT rate vectors can be efficiently computed using Procedure 2. Moreover, we proved that the  $\alpha$ -fair tradeoff policy achieves the optimal EJT for the special case of linear networks with uniform capacity. However, we also illustrated that the  $\alpha$ -fair tradeoff policy does not achieves the optimal EJT for general networks.
- We proposed a novel ICIC scheme that runs in polynomial time and finds a near-optimum dynamic multi-cell resource allocations that maximize a weighted sum-rate in the entire network in a distributed manner. The proposed algorithm can be used with any AMC scheme, including discrete-rate AMC schemes. The proposed scheme is developed using primal-decomposition method, which is utilized to decompose the problem into a master problem and multiple subproblems. We showed that each subproblem has a special network flow structure. By exploiting such network structure, each sub-problem is solved using network-based optimization methods, which have significantly reduced complexity in comparison with general-purpose convex or linear optimization methods.
- Through extensive simulation results, we showed that the proposed ICIC scheme

achieves high gain in aggregate throughput, Jain's index, cell-edge throughput, and outage probability, as compared to reuse-1, reuse-3, partial frequency reuse, dynamic fractional frequency reuse, and optimum fractional frequency reuse.

- We demonstrated that the proposed ICIC scheme can significantly enhance the EJT for nonuniform user distribution.
- We proposed ICIC schemes for HetNets. Unlike most of the existing schemes, we demonstrated that it is possible to improve the performance of macro-UTs and pico-UTs, simultaneously, at the expense of fast message exchange. We also proposed low-complexity variants that provide a good compromise between complexity and gains in performance.

## 6.2 Future Work

The work presented in this thesis raises some interesting questions for future research.

- In this research, scarce radio resources are allocated in a way that optimizes objective measures based on delivered rates and can be extended to accommodate for transmission delays. However, this approach does not necessarily optimize the user's experience as it does not distinguish traffic that has less importance from the users' perspective (e.g., application updates) and traffic that has more importance (e.g., driving navigation). Recently, context-aware resource allocation is introduced as a recent paradigm shift that aims at optimizing users' experience. This is accomplished by making resource allocation decisions based on users' context information such as their locations, preferences, currently running applications, buffer state, and/or urgency of communication [104]. By exploiting the context information, more efficient resource allocation can be made

without sacrificing users' experience. Towards this end, we envision that developing context-aware radio resource allocation schemes can be far more superior to what is currently available.

There are a number of challenges associated with translating context-aware resource allocation into a comprehensive optimization framework. The main challenge is to convert the context information that governs users' experience, which are usually subjective quantities, into appropriate objective measures that are indicative of users' experience and suitable for optimization. Another fundamental challenge is to incorporate the notion of fairness between users in this context, since different users perceive their experience differently. Another challenge is to devise low-complexity distributed interference/context-aware resource allocation through inter-cell coordination that solve the resulting optimization problem.

- In this research, we developed optimal EJT schedulers and devised distributed ICIC schemes for full-buffer traffic. An important extension of this work is to study the case of different traffic classes with different delay requirements and different burstiness levels. In such a case, Jain's fairness index is not suitable and hence it is interesting to devise a fundamentally different efficiency-fairness tradeoff framework.
- We showed that the  $\alpha$ -fair policy does not achieve the optimal EJT in general. However, we also identified two special cases where the  $\alpha$ -fair policy achieves the optimal EJT, namely, the case of two users and the case of congestion control in linear network with uniform capacity. This raises the following question: Under what conditions does the  $\alpha$ -fair policy achieve the optimal EJT? Answering this question will allow us to identify broader class of resource allocation problems where the  $\alpha$ -fair policy achieves the optimal EJT.

- We illustrated that one of the advantages of using Jain’s index as a fairness measure in efficiency-fairness tradeoff optimization is that it improves the median rate and the satisfaction of the users as compared to the  $\alpha$ -fair and MTMR tradeoff policies. However, we have also observed that optimizing with respect to Jain’s index can degrade the rate achieved by cell-edge users. In order to reduce the effect of such degradation, a minimum rate constraint is imposed. While this can potentially rectify the problem, it requires parameter-tuning for the minimum rate bound. A promising approach for future research is to modify Jain’s fairness index such that it implicitly accounts for cell-edge users.
- In this research, we assumed that all BSs have backhaul connections that can carry all the traffic generated by the UTs. While this assumption holds true for macro-BSs, it may not hold in next generation cellular HetNets where pico-BSs and femto-BSs are massively deployed without prior planning. Indeed, pico-BSs and femto-BSs are expected to have backhaul connections with limited capacities. For instance, femto-BSs can be deployed by consumers and they may use DSL connections. An interesting extension to this research is to develop efficiency-fairness framework for interference-limited HetNet with capacity-limited backhaul connections.
- In HetNets, we demonstrated that it is possible to simultaneously improve the performance of both macro-UTs and pico-UTs at the expense of fast message exchange between BSs. In massive deployment of BSs, such fast exchange may not be feasible between all BSs. Thus, we proposed low complexity variants of the proposed ICIC schemes where the message exchange frequency can be adjusted to suit the practical limitations. An interesting research problem is to theoretically bound the performance degradation caused by limited (or no) message exchange as compared to fast message exchange.

## Appendix A

# Published Papers, Submitted Manuscripts, and Patent Applications

The following is a list of the publications as well as the patent applications produced during the PhD program. I am the first author and the main contributor of these publications and patent applications. Other publications in which I am not the first author and/or my role was mainly supporting or supervising were not included in this list.

### A.1 Patents

- **Akram Bin Sediq**, Ramy H. Gohary, Halim Yanıkömeroğlu, Gamini Senarath, and Ho Ting Cheng, “System and method for Jain scheduling in wireless networks,” U.S. Provisional Patent Application 61/720,230, filed by Huawei-Canada, October 2012.
- **Akram Bin Sediq**, Rainer Schoenen, Halim Yanıkömeroğlu, Gamini Senarath, Zhijun Chao, Ho Ting Cheng, and Peiying Zhu, “Inter-cell interference coordination for wireless communication systems,” U.S. Patent Application 13/438,624, filed by Huawei-Canada, April 2012.

- **Akram Bin Sediq**, Jietao Zhang, Petar Djukic, and Halim Yanıkömeroğlu, “Data transmission method, system, intermediate node and source node in cooperative relay network,” International Patent Application PCT/CN2011/071 286, filed by Huawei-China, April 2010<sup>1</sup>.

## A.2 Journal Papers

- **Akram Bin Sediq**, Ramy H. Gohary, Rainer Schoenen, and Halim Yanıkömeroğlu, “Optimal tradeoff between sum-rate efficiency and Jain’s fairness index in resource allocation,” *IEEE Transactions on Wireless Communications*, vol. 12, no. 7, pp. 3496–3509, July 2013.
- **Akram Bin Sediq**, Rainer Schoenen, Halim Yanıkömeroğlu, and Gamini Senarath, “Optimized distributed inter-cell interference coordination scheme using projected subgradient and network flow optimization,” under review in *IEEE Transactions on Communications* (submission: 22 March 2013, 1st results: 10 June 2013).
- **Akram Bin Sediq** and Halim Yanıkömeroğlu, “Selection combining of signals with different modulation levels in Nakagami-m fading,” *IEEE Communications Letters*, vol. 16, no. 5, pp. 752–755, May 2012<sup>2</sup>.
- **Akram Bin Sediq**, Petar Djukic, Halim Yanıkömeroğlu, and Jietao Zhang, “Optimized non-uniform constellation rearrangement for cooperative relaying,” *IEEE Transactions on Vehicular Technology*, vol. 60, no. 5, pp. 2340–2347, June 2011<sup>1</sup>.

---

<sup>1</sup>This work was conducted during the first two years of the PhD program. Due to coherence issues, this work is not included in the PhD thesis.

<sup>2</sup>This work was conducted during the third year of the PhD program. Due to coherence issues, this work is not included in the PhD thesis.

### A.3 Conference Papers

- **Akram Bin Sediq**, Ramy H. Gohary, and Halim Yanıkömeroğlu, “Optimal tradeoff between efficiency and Jains fairness index in resource allocation,” *IEEE International Symposium on Personal, Indoor and Mobile Communications (PIMRC 2012)*, 9–12 September 2012, Sydney, Australia.
- **Akram Bin Sediq**, Rainer Schoenen, Halim Yanıkömeroğlu, Gamini Senarath, and Zhijun Chao, “A novel distributed inter-cell interference coordination scheme based on projected subgradient and network flow optimization,” *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC 2011)*, 11–14 September 2011, Toronto, Canada.
- **Akram Bin Sediq**, Petar Djukic, Halim Yanıkömeroğlu, and Jietao Zhang, “Generalized constellation rearrangement in cooperative relaying,” *IEEE Vehicular Technology Conference (VTC2010-Spring)*, 16–19 May 2010, Taipei, Taiwan<sup>1</sup>.
- **Akram Bin Sediq**, Petar Djukic, Halim Yanıkömeroğlu, and Jietao Zhang, “Near-optimal non-uniform constellation rearrangement for cooperative relaying,” *25th Biennial Symposium on Communications (QBSC 2010)*, 12–14 May 2010, Queens University, Kingston, Ontario, Canada<sup>1</sup>.

## List of References

- [1] “Cisco visual networking index: Global mobile data traffic forecast update, 2012-2017,” White Paper, Cisco Systems Inc., Feb. 2013. [Online]. Available: [http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white\\_paper\\_c11-520862.pdf](http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-520862.pdf)
- [2] R. Jain, D. Chiu, and W. Hawe, “A quantitative measure of fairness and discrimination for resource allocation in shared systems,” Digital Equipment Corporation, DEC-TR-301, Tech. Rep., 1984. [Online]. Available: <http://www1.cse.wustl.edu/~jain/papers/ftp/fairness.pdf>.
- [3] J. Mo and J. Walrand, “Fair end-to-end window-based congestion control,” *IEEE/ACM Trans. Netw.*, vol. 8, no. 5, pp. 556–567, Oct. 2000.
- [4] R. Ahuja, T. Magnanti, and J. Orlin, *Network Flows: Theory, Algorithms, and Applications*. Prentice Hall, 1993.
- [5] T. Lan, D. Kao, M. Chiang, and A. Sabharwal, “An axiomatic theory of fairness in network resource allocation,” in *Proc. IEEE Int. Conf. Comp. Commun. (INFOCOM)*, San Diego, California, Mar. 2010.
- [6] H. T. Cheng and W. Zhuang, “An optimization framework for balancing throughput and fairness in wireless networks with QoS support,” *IEEE Trans. Wireless Commun.*, vol. 7, no. 2, pp. 584–593, Feb. 2008.
- [7] R. Schoenen, A. Bin Sediq, H. Yanikomeroglu, G. Senarath, and Z. Chao, “Fairness analysis in cellular networks using stochastic Petri nets,” in *Proc. IEEE Int. Symp. on Pers., Indoor and Mobile Radio Commun. (PIMRC)*, Toronto, Canada, Sept. 2011.
- [8] R. Schoenen, A. Bin Sediq, H. Yanikomeroglu, G. Senarath, Z. Chao, and H. T. Cheng, “Spectral efficiency and fairness tradeoffs in cellular networks

- with realtime+nonrealtime traffic mix using stochastic Petri nets,” in *Proc. IEEE Veh. Tech. Conf. (VTC-Fall)*, Quebec City, Canada, Sept. 2012.
- [9] D. Bertsekas and R. Gallager, *Data Networks*, 2nd ed. Prentice Hall, 1992.
- [10] A. Odlyzko, “Network neutrality, search neutrality, and the never-ending conflict between efficiency and fairness in markets,” *Rev. Netw. Economics*, vol. 8, no. 1, pp. 40–60, Mar. 2009.
- [11] M. Zukerman, L. Tan, H. Wang, and I. Ouveysi, “Efficiency-fairness tradeoff in telecommunications networks,” *IEEE Commun. Lett.*, vol. 9, no. 7, pp. 643–645, July 2005.
- [12] Z. Shen, J. G. Andrews, and B. L. Evans, “Adaptive resource allocation in multiuser OFDM systems with proportional rate constraints,” *IEEE Trans. Wireless Commun.*, vol. 4, no. 6, pp. 2726–2737, Nov. 2005.
- [13] F. P. Kelly, A. K. Maulloo, and D. K. H. Tan, “Rate control for communication networks: Shadow prices, proportional fairness and stability,” *J. Oper. Res. Soc.*, vol. 49, no. 3, pp. 237–252, 1998.
- [14] C. Wengerter, J. Ohlhorst, and A. von Elbwart, “Fairness and throughput analysis for generalized proportional fair frequency scheduling in OFDMA,” in *Proc. IEEE Veh. Tech. Conf. (VTC-Spring)*, Stockholm, Sweden, May 2005.
- [15] C. Joe-Wong, S. Sen, T. Lan, and M. Chiang, “Multi-resource allocation: Fairness-efficiency tradeoffs in a unifying framework,” in *Proc. IEEE Int. Conf. Comp. Commun. (INFOCOM)*, Orlando, Florida, Mar. 2012.
- [16] A. Ibing and H. Boche, “Fairness vs. efficiency: Comparison of game theoretic criteria for OFDMA scheduling,” in *Proc. Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, California, Nov. 2007.
- [17] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.
- [18] T. Lan and M. Chiang, “An axiomatic theory of fairness in resource allocation,” Tech. Rep., 2009. [Online]. Available: [www.princeton.edu/~chiangm/fairness.pdf](http://www.princeton.edu/~chiangm/fairness.pdf)
- [19] L. Kudryavtsev, “Homogeneous function,” [http://www.encyclopediaofmath.org/index.php?title=Homogeneous\\_function&oldid=11366](http://www.encyclopediaofmath.org/index.php?title=Homogeneous_function&oldid=11366).

- [20] S. Schwarz, C. Mehlhruer, and M. Rupp, “Throughput maximizing multiuser scheduling with adjustable fairness,” in *Proc. IEEE Int. Conf. Commun. (ICC)*, June 2011, pp. 1–5.
- [21] A. Bin Sediq, R. H. Gohary, and H. Yanikomeroglu, “Optimal tradeoff between efficiency and Jain’s fairness index in resource allocation,” in *Proc. IEEE Int. Symp. on Pers., Indoor and Mobile Radio Commun. (PIMRC)*, Sydney, Australia, Sept. 2012.
- [22] B. Radunovic and J.-Y. L. Boudec, “Rate performance objectives of multihop wireless networks,” *IEEE Trans. Mobile Comput.*, vol. 3, no. 4, pp. 334–349, Oct. 2004.
- [23] G. Song and Y. Li, “Cross-layer optimization for OFDM wireless networks—part II: Algorithm development,” *IEEE Trans. Wireless Commun.*, vol. 4, no. 2, pp. 625–634, Mar. 2005.
- [24] A. Tang, J. Wang, and S. Low, “Counter-intuitive throughput behaviors in networks under end-to-end control,” *IEEE/ACM Trans. Netw.*, vol. 14, no. 2, pp. 355–368, Apr. 2006.
- [25] T. Bonald and L. Massoulié, “Impact of fairness on internet performance,” in *Proc. ACM SIGMETRICS*, Cambridge, Massachusetts, June 2001.
- [26] C. Y. Wong, R. S. Cheng, K. Ben Letaief, and R. D. Murch, “Multiuser OFDM with adaptive subcarrier, bit, and power allocation,” *IEEE J. Sel. Areas Commun.*, vol. 17, no. 10, pp. 1747–1758, Oct. 1999.
- [27] C. Guo, M. Sheng, Y. Zhang, and X. Wang, “A Jain’s index perspective on  $\alpha$ -fairness resource allocation over slow fading channels,” *IEEE Commun. Lett.*, vol. 17, no. 4, pp. 705–708, 2013.
- [28] P. Viswanath, D. Tse, and R. Laroia, “Opportunistic beamforming using dumb antennas,” *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1277–1294, June 2002.
- [29] E. Dahlman, S. Parkvall, J. Skold, and P. Beming, *3G Evolution: HSPA and LTE for Mobile Broadband*, 2nd ed. Academic Press, 2008.
- [30] E. Biglieri, J. Proakis, and S. Shamai (Shitz), “Fading channels: Information-theoretic and communication aspects,” *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2619–2692, Oct. 1998.

- [31] A. L. Stolyar, “On the asymptotic optimality of the gradient scheduling algorithm for multiuser throughput allocation,” *Oper. Res.*, vol. 53, no. 1, pp. 12–25, Jan. 2005.
- [32] H. Kushner and P. Whiting, “Convergence of proportional-fair sharing algorithms under general conditions,” *IEEE Trans. Wireless Commun.*, vol. 3, no. 4, pp. 1250–1259, July 2004.
- [33] M. Andrews, L. Qian, and A. Stolyar, “Optimal utility based multi-user throughput allocation subject to throughput constraints,” in *Proc. IEEE Int. Conf. Comp. Commun. (INFOCOM)*, Miami, Florida, Mar. 2005.
- [34] ITU, Report ITU-R M.2135-1, “Guidelines for evaluation of radio interface technologies for IMT-Advanced,” ITU, Tech. Rep., Dec. 2009. [Online]. Available: <http://www.itu.int/pub/R-REP-M.2135-1-2009>.
- [35] J. Nystrom, “Software implementation of IMT.Eval channel model,” CELTIC/CP5-026 Project WINNER+ Doc. 5D/478-E, Tech. Rep., July 2009. [Online]. Available: <http://www.itu.int/oth/R0A06000022/en>.
- [36] R. Schoenen and C. Teijeiro, “System level performance evaluation of LTE with MIMO and relays in reuse-1 IMT-Advanced scenarios,” in *Proc. IEEE WiCom*, Chengdu, China, Sept. 2010.
- [37] “Calibration for IMT-Advanced evaluations,” CELTIC/CP5-026 Project WINNER+, Tech. Rep., May 2010. [Online]. Available: <http://projects.celtic-initiative.org/winner+/WINNER+EvaluationGroup.html>.
- [38] IEEE 802.16 Broadband Wireless Access Working Group, “IEEE 802.16m evaluation methodology doc. (EMD),” Tech. Rep., Jan. 2009. [Online]. Available: [http://www.ieee802.org/16/tgm/docs/80216m-08\\\_004r2.pdf](http://www.ieee802.org/16/tgm/docs/80216m-08\_004r2.pdf).
- [39] “LTE Advanced: Heterogeneous networks,” White Paper, Qualcomm Incorporated, Jan. 2011. [Online]. Available: <http://www.qualcomm.com/media/documents/files/qualcomm-research-lte-heterogeneous-networks.pdf>
- [40] T. S. Rappaport, *Wireless Communications: Principles and Practice*. Prentice-Hall, 1996.
- [41] 3GPP Project Doc. R1-050 507, “Soft frequency reuse scheme for UTRAN LTE,” Tech. Rep., May 2005.

- [42] M. Sternad, T. Ottosson, A. Ahlen, and A. Svensson, "Attaining both coverage and high spectral efficiency with adaptive OFDM downlinks," in *Proc. IEEE Veh. Tech. Conf. (VTC-Fall)*, Orlando, Florida, Oct. 2003.
- [43] Y. Xiang, J. Luo, and C. Hartmann, "Inter-cell interference mitigation through flexible resource reuse in OFDMA based communication," in *Proc. 13th European Wireless Conf.*, Paris, France, Apr. 2007.
- [44] A. Simonsson, "Frequency reuse and intercell interference co-ordination in E-UTRA," in *Proc. IEEE Veh. Tech. Conf. (VTC-Spring)*, Dublin, Ireland, Apr. 2007.
- [45] 3GPP Project Doc. R1-072 762, "Further discussion on adaptive fractional frequency reuse," Tech. Rep., June 2007.
- [46] S. H. Ali and V. C. M. Leung, "Dynamic frequency allocation in fractional frequency reused OFDMA networks," *IEEE Trans. Wireless Commun.*, vol. 8, no. 12, pp. 4286–4295, Aug. 2009.
- [47] X. Zhang, C. He, L. Jiang, and J. Xu, "Inter-cell interference coordination based on softer frequency reuse in OFDMA cellular systems," in *Proc. Int. Conf. on Neural Net. and Signal Process. (ICNNSP)*, Zhenjiang, China, June 2008.
- [48] X. Mao, A. Maaref, and K. H. Teo, "Adaptive soft frequency reuse for inter-cell interference coordination in SC-FDMA based 3GPP LTE uplinks," in *Proc. IEEE Glob. Commun. Conf. (Globecom)*, New Orleans, LA, Dec. 2008.
- [49] A. Stolyar and H. Viswanathan, "Self-organizing dynamic fractional frequency reuse for best-effort traffic through distributed inter-cell coordination," in *Proc. IEEE Int. Conf. Comp. Commun. (INFOCOM)*, Rio de Janeiro, Brazil, Apr. 2009.
- [50] G. Li and H. Liu, "Downlink radio resource allocation for multi-cell OFDMA system," *IEEE Trans. Wireless Commun.*, vol. 5, no. 12, pp. 3451–3459, Dec. 2006.
- [51] R. Y. Chang, Z. Tao, J. Zhang, and C. C. J. Kuo, "Multicell OFDMA downlink resource allocation using a graphic framework," *IEEE Trans. Veh. Technol.*, vol. 58, no. 12, pp. 3494–3507, Sept. 2009.
- [52] J. Ellenbeck, C. Hartmann, and L. Berlemann, "Decentralized inter-cell interference coordination by autonomous spectral reuse decisions," in *Proc. 14th European Wireless Conf.*, Prague, Czech Republic, June 2008.

- [53] M. Rahman and H. Yanikomeroglu, “Enhancing cell-edge performance: A downlink dynamic interference avoidance scheme with inter-cell coordination,” *IEEE Trans. Wireless Commun.*, vol. 9, no. 4, pp. 1414–1425, Apr. 2010.
- [54] Z. Han, Farrokhi, Z. Ji, and K. Liu, “Capacity optimization using subspace method over multicell OFDMA networks,” in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Atlanta, Georgia, Mar. 2004.
- [55] M. Assaad, “Optimal fractional frequency reuse (FFR) in multicellular OFDMA system,” in *Proc. IEEE Veh. Tech. Conf. (VTC-Fall)*, Marina Bay, Singapore, Sept. 2008.
- [56] L. Chen and D. Yuan, “Beyond conventional fractional frequency reuse for networks with irregular cell layout: An optimization approach and performance evaluation,” in *The 5th Annu. Wireless Internet Conf. (WICON)*, Singapore, Singapore, Mar. 2010.
- [57] F. Xiangning, C. Si, and Z. Xiaodong, “An inter-cell interference coordination technique based on users ratio and multi-level frequency allocations,” in *Int. Conf. Wireless Commun., Netw. and Mobile Comput.*, Shanghai, China, Sept. 2007.
- [58] I. Fraimis, V. Papoutsis, and S. Kotsopoulos, “A decentralized subchannel allocation scheme with inter-cell interference coordination (ICIC) for multi-cell OFDMA systems,” in *Proc. IEEE Glob. Commun. Conf. (Globecom)*, Miami, Florida, Dec. 2010.
- [59] K. Dong, H. Tian, X. Li, and Q. Sun, “A distributed inter-cell interference coordination scheme in downlink multicell OFDMA systems,” in *Proc. 7th IEEE Consumer Commun. and Networking Conf. (CCNC)*, Las Vegas, Nevada, Jan. 2010.
- [60] D. Kimura, Y. Harada, and H. Seki, “De-centralized dynamic ICIC using X2 interfaces for downlink LTE systems,” in *Proc. IEEE Veh. Tech. Conf. (VTC-Spring)*, Budapest, Hungary, May 2011.
- [61] I. G. Fraimis, V. D. Papoutsis, and S. A. Kotsopoulos, “A distributed radio resource allocation algorithm with interference coordination for multi-cell OFDMA systems,” in *Proc. IEEE Int. Symp. on Pers., Indoor and Mobile Radio Commun. (PIMRC)*, Istanbul, Turkey, Sept. 2010.

- [62] A. Damnjanovic, J. Montojo, J. Cho, H. Ji, J. Yang, and P. Zong, “UE’s role in LTE advanced heterogeneous networks,” *IEEE Commun. Mag.*, vol. 50, no. 2, pp. 164 – 176, Feb. 2012.
- [63] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge University Press, 2005.
- [64] C. Y. Wong, R. S. Cheng, K. Ben Letaief, and R. D. Murch, “Multiuser OFDM with adaptive subcarrier, bit, and power allocation,” *IEEE J. Sel. Areas Commun.*, vol. 17, no. 10, pp. 1747–1758, Oct. 1999.
- [65] W. Rhee and J. M. Cioffi, “Increase in capacity of multiuser OFDM system using dynamic subchannel allocation,” in *Proc. IEEE Veh. Technol. Conf.*, Tokyo, Japan, May 2000.
- [66] Z. Shen, J. G. Andrews, and B. L. Evans, “Optimal power allocation in multiuser OFDM systems,” in *Proc. IEEE Glob. Commun. Conf. (Globecom)*, San Francisco, Dec. 2003.
- [67] J. Jang and K. B. Lee, “Transmit power adaptation for multiuser OFDM systems,” *IEEE J. Sel. Areas Commun.*, vol. 21, no. 2, pp. 171–178, Feb. 2003.
- [68] M. Chiang, “Balancing transport and physical layers in wireless multihop networks: Jointly optimal congestion control and power control,” *IEEE J. Sel. Areas Commun.*, vol. 23, no. 1, pp. 104–116, Jan. 2005.
- [69] M. Chiang, C. W. Tan, D. P. Palomar, D. O’Neill, and D. Julian, “Power control by geometric programming,” *IEEE Trans. Wireless Commun.*, vol. 6, no. 7, pp. 2640–2651, July 2007.
- [70] C. W. Tan, M. Chiang, and R. Srikant, “Fast algorithms and performance bounds for sum rate maximization in wireless networks,” *IEEE/ACM Trans. Netw.*, accepted June 30, 2012.
- [71] R. D. Yates, “A framework for uplink power control in cellular radio systems,” *IEEE J. Sel. Areas Commun.*, vol. 13, no. 7, pp. 1341–1347, Sept. 1995.
- [72] M. Chiang, P. Hande, T. Lan, and C. W. Tan, *Power Control in Wireless Cellular Networks*. Now Publishers Inc, 2008.
- [73] A. Gjendemsjo, D. Gesbert, G. E. Oien, and S. G. Kiani, “Binary power control for sum rate maximization over multiple interfering links,” *IEEE Trans. Wireless Commun.*, vol. 7, no. 8, pp. 3164–3173, Aug. 2008.

- [74] L. P. Qian and Y. J. A. Zhang, “Optimal Throughput-oriented Power Control by Linear Multiplicative Fractional Programming,” in *Proc. IEEE Int. Conf. Commun. (ICC)*, Beijing, China, May 2008.
- [75] D. Gesbert, S. Kiani, A. Gjendemsj, and G. Oien, “Adaptation, coordination, and distributed resource allocation in interference-limited wireless networks,” *Proc. IEEE*, vol. 95, no. 12, pp. 2393 – 2409, Dec. 2007.
- [76] W. Yu and R. Lui, “Dual methods for nonconvex spectrum optimization of multicarrier systems,” *IEEE Trans. Commun.*, vol. 54, no. 7, pp. 1310–1322, July 2006.
- [77] R. Cendrillon, W. Yu, M. Moonen, J. Verlinden, and T. Bostoen, “Optimal multiuser spectrum balancing for digital subscriber lines,” *IEEE Trans. Commun.*, vol. 54, no. 5, pp. 922–933, May 2006.
- [78] L. Venturino, N. Prasad, and X. Wang, “Coordinated scheduling and power allocation in downlink multicell OFDMA networks,” *IEEE Trans. Veh. Technol.*, vol. 58, no. 6, pp. 2835–2848, July 2009.
- [79] Z.-Q. T. Luo and S. Zhang, “Dynamic spectrum management: Complexity and duality,” *IEEE J. Select. Topics Signal Processing*, vol. 2, no. 1, pp. 57–73, Feb. 2008.
- [80] L. P. Qian, Y. J. A. Zhang, and J. Huang, “Mapel: Achieving global optimality for a non-convex wireless power control problem,” *IEEE Trans. Wireless Commun.*, vol. 8, no. 3, pp. 1553–1563, Mar. 2009.
- [81] C. W. Tan, S. Friedland, and S. H. Low, “Spectrum management in multiuser cognitive wireless networks: Optimality and algorithm,” *IEEE J. Sel. Areas Commun.*, vol. 29, no. 2, pp. 421–430, Feb. 2011.
- [82] M. Rossi, A. Tulino, O. Simeone, and A. Haimovich, “Non-convex utility maximization in gaussian MISO broadcast and interference channels,” in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, Apr. 2011.
- [83] E. Bjornson, G. Zheng, M. Bengtsson, and B. Ottersten, “Robust monotonic optimization framework for multicell MISO systems,” *IEEE Trans. Signal Process.*, vol. 60, no. 5, pp. 2508–2523, May 2012.

- [84] A. Gjendemsjo, D. Gesbert, G. E. Oien., and S. G. Kiani, “Optimal Power Allocation and Scheduling for Two-Cell Capacity Maximization,” in *4th International Symp, on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks*, Boston, Massachusetts, Apr. 2006.
- [85] M. Ebrahimi, M. A. Maddah-Ali, and A. K. Khandani, “Power allocation and asymptotic achievable sum-rates in single-hop wireless networks,” in *40th Annu. Conf. on Inform. Sci. and Syst.*, Princeton, NJ, Mar. 2006, pp. 498–503.
- [86] S. R. Bhaskaran, S. V. Hanly, N. Badruddin, and J. S. Evans, “Maximizing the sum rate in symmetric networks of interfering links,” *IEEE Trans. Inf. Theory*, vol. 56, no. 9, pp. 4471–4487, Sept. 2010.
- [87] H. Inaltekin and S. V. Hanly, “Optimality of binary powercontrol in a single cell via majorization,” [arXiv:1105.5861 \[cs.IT\]](https://arxiv.org/abs/1105.5861), May 2011.
- [88] C. S. Chen, K. W. Shum, and C. W. Sung, “Round-robin power control for the weighted sum rate maximisation of wireless networks over multiple interfering links,” *Europ. Trans. Telecommun.*, vol. 22, no. 8, pp. 458–470, Dec. 2011.
- [89] D. P. Bertsekas, *Nonlinear Programming*, 2nd ed. Athena Scientific, 2008.
- [90] S. Boyd, L. Xiao, A. Mutapcic, and J. Mattingley, “Notes on decomposition methods,” Lecture notes for EE364B, Stanford University, Feb. 2007.
- [91] S. Boyd and A. Mutapcic, “Subgradient methods,” Lecture notes for EE364b, Stanford University, Jan. 2007.
- [92] A. Bin Sediq, R. Schoenen, H. Yanikomeroglu, G. Senarath, and Z. Chao, “A novel distributed inter-cell interference coordination scheme based on projected subgradient and network flow optimizations,” in *Proc. IEEE Int. Symp. on Pers., Indoor and Mobile Radio Commun. (PIMRC)*, Toronto, Canada, Sept. 2011.
- [93] J. Andrews, “Seven ways that hetnets are a cellular paradigm shift,” *IEEE Commun. Mag.*, vol. 51, no. 3, pp. 136–144, 2013.
- [94] E. Dahlman, S. Parkvall, and J. Skold, *4G:LTE/LTE-Advanced for Mobile Broadband*, 1st ed. Academic Press, 2011.
- [95] A. Ghosh, R. Ratasuk, B. Mondal, N. Mangalvedhe, and T. Thomas, “Lte-advanced: next-generation wireless broadband technology,” *IEEE Wireless Commun.*, vol. 17, no. 3, pp. 10–22, June 2010.

- [96] V. Pauli, J. D. Naranjo, and E. Seidel, "Heterogeneous LTE networks and inter-cell interference coordination," White Paper, Nomor Research, Dec. 2010. [Online]. Available: [http://lteportal.com/Files/MarketSpace/Download/447\\_1-14nomor.pdf](http://lteportal.com/Files/MarketSpace/Download/447_1-14nomor.pdf)
- [97] A. Ghosh, N. Mangalvedhe, R. Ratasuk, B. Mondal, M. C. E. Visotsky, T. A. Thomas, J. G. Andrews, P. Xia, H. S. Jo, H. S. Dhillon, and T. D. Novlan, "Heterogeneous cellular networks: From theory to practice," *IEEE Commun. Mag.*, vol. 50, no. 6, pp. 54–64, June 2012.
- [98] L. Lindbom, R. Love, S. Krishnamurthy, C. Yao, N. Miki, and V. Chandrasekhar, "Enhanced inter-cell interference coordination for heterogeneous networks in LTE-Advanced: A survey," *CoRR*, vol. abs/1112.1344, 2011.
- [99] 3GPP TR 36.814-900, "3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Further Advancements for E-UTRA Physical Layer Aspects (Release 9)," Tech. Rep., Mar. 2010.
- [100] D. Lopez-Perez, A. Valcarce, G. de la Roche, and J. Zhang, "OFDMA femto-cells: A roadmap on interference avoidance," *IEEE Commun. Mag.*, vol. 47, no. 9, pp. 41–48, Sept. 2009.
- [101] D. Lopez-Perez, I. Guvenc, G. De la Roche, M. Kountouris, T. Quek, and J. Zhang, "Enhanced intercell interference coordination challenges in heterogeneous networks," *IEEE Wireless Commun.*, vol. 18, no. 3, pp. 22–30, 2011.
- [102] K. Pedersen, Y. Wang, B. Soret, and F. Frederiksen, "eICIC functionality and performance for LTE hetnet co-channel deployments," in *Proc. IEEE Veh. Tech. Conf. (VTC-Fall)*, Qubec City, Canada, 2012.
- [103] 3GPP TS 36.300 V10.3.0, "Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall description; Stage 2 (Release 10)." 3GPP, Tech. Rep.
- [104] M. Proebster, M. Kaschub, T. Werthmann, and S. Valentin, "Context-aware resource allocation for cellular wireless networks," *EURASIP Wireless Commun. Netw.*, vol. 2012, July 2012.