

Radio Resource Management in OFDMA-based Cellular Relay Networks

Mohamed Salem, B.Sc., M.Sc.

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 Engineering

Ottawa–Carleton Institute for Electrical and Computer Engineering
Department of Systems and Computer Engineering
Carleton University, Ottawa, Ontario, Canada

© 2011 Mohamed Salem

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

Radio Resource Management in OFDMA-based Cellular Relay Networks

Submitted by Mohamed Adel M. Rashad Salem
in partial fulfilment of the requirements for the degree of
Doctor of Philosophy in Electrical Engineering

Chair, Department of Systems and Computer Engineering
Professor Howard M. Schwartz

Thesis Supervisor
Professor Halim Yanikömeroglu

Thesis Supervisor
Professor David Falconer

External Examiner
Professor Ekram Hossain
Dept. of Electrical and Computer Eng., U Manitoba

Carleton University

© January 2011

Abstract

The main objective of next-generation wireless networks is to accommodate the increasing user demand and to achieve a ubiquitous high-data-rate coverage so that mobile broadband services comparable to those of the wirelines are realized in a cost-efficient manner. This ambitious objective however faces several technical challenges in the conventional cellular architecture, e.g., the large pathloss. Wireless multihop relaying is therefore among the envisioned solutions; for that relays- with much less functionality and cost than base stations- can extend coverage, overcome shadowing through routing, and/or improve detection in a cooperative manner. Hence, a future network comprising various forms of dedicated wireless relays is envisaged in many wireless standardization bodies and forums which have adopted orthogonal frequency division multiple access (OFDMA) as the prospective air interface.

The synergy of multihop relaying and OFDMA techniques offers a very rich set of opportunities but renders an unsuitable environment for static resource planning. This is due to the increased system dynamics, the portion of resources invested in operating the wireless relays, and the co-channel interference associated with the inevitable aggressive reuse of the scarce licensed spectrum. Therefore, intelligent radio resource management (RRM) schemes are required to combat the interference and to operate the relays in a dynamic and opportunistic manner. It is not immediately clear though how to perform RRM in such a complex environment. As such, RRM could be the main obstacle confronting the deployment and operation of future networks.

On these grounds, our objective is to devise intelligent RRM schemes that dynamically optimize the resource allocation in future networks in multiple dimensions; frequency, time, space, power, and relay route, so that bandwidth is efficiently utilized while a reliable and ubiquitous service is achieved regardless of users' locations and channel conditions. Towards that end, mathematical optimization and dynamic programming tools are employed along with novel opportunistic medium access techniques, in a centralized as well as a distributed manner. Various forms of relays of different characteristics and functionalities, such as the fixed and nomadic relays, are incorporated. In addition, algorithms with low computational complexity and signaling overhead are devised for practical implementation.

Acknowledgement

In the name of God the Most Gracious, the Most Merciful. To Him we belong and all the praise and thanks are due. By only His will our efforts could ever be fruitful and our plans could ever be seen through. No matter how much knowledge we acquire, He remains the utmost knowing and before His name, none of the following deservings may show.

My deep and sincere thanks, gratitude and respect are due to my academic supervisor and research mentor, Prof. Halim Yanıkömeroğlu. No doubt, I would attribute the success of this program to his on-going guidance and phenomenal mentorship that was never limited to the highly professional academic and technical supervision. I must also acknowledge his persistent encouragement and inspiring enthusiasm to conduct high quality research publishable in the first tier journals and flagship international conferences. He has been more of a friend or an older brother to me than a supervisor. In a nutshell, Prof. Yanıkömeroğlu has made the usually cumbersome and stressful PhD course of study an enjoyable and unforgettable experience for me.

My appreciation and gratitude are due as well to my thesis co-supervisor, Prof. David Falconer whose invaluable comments and professional insights have substantially enhanced the quality of this work, not to mention his persistent and strong support in grant and award applications throughout the program.

I would like also to thank my project manager, my co-author, and my dear friend, Dr. Abdulkareem Adinoyi, for all his sincere efforts, sleepless nights and diligent work

he has put into that project to make it such a success. He has indeed provided me with an ideal research environment that was full of appreciation, brotherhood, and productive teamwork.

Sincere thanks to all of my colleagues and fellow researchers who helped me see this dream coming true, especially Mahmudur Rahman, whose technical tips and experience have expedited the warming phase of my research; and also Sebastian Szyzkowicz who have kindly helped me shape this thesis into its current format.

I would like also to thank our industrial collaborator, Samsung Electronics, SAIT, Korea, for funding this program, providing enlightening discussions, and filing three patents out of this work.

On the personal side, special thanks and gratitude to my beloved parents for their continuous support and encouragement throughout my career. I just can not thank them enough for every single moment and effort they invested on my path to earn this degree. May the success of this program be the sweet fruit of their long-term strive. I would like also to express my appreciation and gratefulness to my beloved wife for her tireless support, understanding, patience, and countless sacrifices since I started this program. There are no words indeed that can be said in return to her loving care. Although he is only one year old, I wouldn't deny as well the great influence my little son had on my enthusiasm and motivation to attain such an honor that would make him proud once he is old enough to read and comprehend these lines. Last but not least, I would like to thank my sister and my brother for their warm encouragement and moral support.

Contents

Abstract	iii
Acknowledgement	v
Contents	vii
List of Figures	xii
List of Tables	xviii
Nomenclature	xix
1 Introduction	1
1.1 The RRM Problem Statement	2
1.2 Publications, Patent filings, and Technical Reports	5
2 Overview on RRM in OFDMA Relay Networks	10
2.1 Activities on Relay-based Architecture among Standardization Bodies, Forums, and Consortiums	11
2.2 Some Basic RRM Tools	13
2.2.1 Scheduling	13
2.2.2 Routing	14
2.2.3 Link Adaptation: Adaptive Modulation and Coding	15

2.2.4	Power Control	15
2.2.5	Cooperative Relaying in OFDMA Networks	16
2.3	Earlier RRM Research in Relay Networks	18
2.3.1	Relaying in Downlink Multicellular Networks	18
2.3.2	Relaying in Uplink Multicellular Networks	21
2.4	RRM in OFDMA Fixed-Relay Networks	24
2.4.1	Centralized RRM Schemes in Single-cell OFDMA Relay Networks	24
2.4.2	Centralized RRM Schemes in Multi-cell OFDMA Relay Networks	30
2.4.3	Distributed RRM Schemes in OFDMA Relay Networks	33
2.5	RRM Opportunities and Challenges in OFDMA Cellular Relay Networks	35
2.5.1	Migration from Conventional Cellular to Relay-Enhanced	35
2.5.1.1	Centralized or Distributed?	35
2.5.1.2	Cell-load Balancing	37
2.5.1.3	Various Forms of Wireless Relays of Different functionalities	37
2.5.1.4	Relay Selection or In-Cell Routing	39
2.5.1.5	CAC and Handover for Network-Load Balancing via Inter-Cell Routing	40
2.5.2	Co-Channel Interference in OFDMA Cellular Relay Networks	41
2.5.3	Restating the Strategy for Improving the Goodput Efficiency	46
2.6	Role of Multi-antenna Systems in the Future Networks	47
2.7	Extending Conventional scheduling algorithms to Cellular Relay Networks	49
2.8	Fairness in OFDMA Relay Networks: An Obligation or A Privilege?	50
2.8.1	Notion of Fairness in this Thesis	52
2.8.2	Fairness Metric Examples	53
2.8.3	Fairness Implementation in RRA Algorithms	55

2.9	Channel Models and Numerical Examples	58
2.10	Conclusions	60
3	Fairness-aware RRM in Downlink OFDMA Cellular Relay Networks	62
3.1	Introduction	62
3.2	Contributions and Relevant Work	64
3.3	System Description and Assumptions	67
3.4	The BS's Joint Routing and Fair Scheduling	68
3.4.1	Mathematical Formulation of the RRA at the BS	70
3.4.2	A Low-complexity Iterative RRA Algorithm	73
3.5	PFS-based RRM in OFDMA Relay Networks	79
3.6	Simulated Network Performance	81
3.6.1	Simulation Models and Parameters	81
3.6.2	Simulation Results and Discussion	81
3.7	Signalling Overhead and Delays	91
3.8	Conclusions	94
4	Fair Resource Allocation Towards Ubiquitous Coverage in OFDMA Cellular Relay Networks with Asymmetric Traffic	95
4.1	Introduction	95
4.2	System Model and Assumptions	97
4.3	Mathematical Formulation of the RRA	100
4.3.1	The Joint Routing and Scheduling for the BS Sub-frame	102
4.3.2	Formulation of Variant-A for the RS Sub-frame	104
4.3.3	Formulation of Variant-B for the RS Sub-frame	106
4.4	Realization of the Schemes Through Low-Complexity Iterative Algorithms	108
4.4.1	Dynamic Routing in the Two-hop Cellular Relay Network	110

4.4.2	The Computational Complexity	114
4.5	Numerical Results	114
4.5.1	The Proposed Scheme vs. the Quasi-FDR scheme and Prior Art	115
4.5.2	Variant-A with symmetric and asymmetric traffic	120
4.6	Implementation Issues and Feedback Overhead	124
4.7	Conclusions	126
5	Integrating Self-Organizing Nomadic Relays into OFDMA Fixed-Relay Cellular Networks	127
5.1	Introduction	127
5.2	Description of the Proposed Systems	132
5.3	The Self-Optimizing WT-based Dynamic Routing	138
5.4	Self-Organizing Nomadic Relay Operations	140
5.4.1	Autonomous Subchannels Acquisition and Selective Relaying .	141
5.4.2	NRS-FRS Cooperation	144
5.5	Radio Resource Allocation at the BS	144
5.5.1	Semi-Centralized Scheme	145
5.5.2	Distributed Scheme	146
5.5.3	Required Feedback Overhead	147
5.6	Simulation Results	151
5.7	Conclusions	160
6	Joint Power and Subchannel Allocation for the Self-Organizing Nomadic Relays in OFDMA-based Cellular Fixed-Relay Networks	161
6.1	Introduction	161
6.2	System Description	164
6.3	NRS Joint Power-Subchannel Allocation Algorithm	164
6.4	System Performance	168

6.4.1	Simulation Parameters and Channel Model	170
6.4.2	Numerical Results and Discussions	170
6.5	Conclusions	177
7	Summary of Conclusions, Contributions, and Possible Extensions	178
7.1	Thesis Conclusions	178
7.2	Thesis Contributions	181
7.3	Extensions to the Centralized Schemes in Chapter 3 and Chapter 4 .	182
7.3.1	Intra-cell Resource Reuse	182
7.3.2	Connection Admission Control for Load Balancing via Inter-cell Routing	183
7.3.3	Partial Feedback: How Much and How Often is Enough? . . .	184
7.4	Extensions to the Decentralized RRM in Chapter 5 and Chapter 6 . .	185
7.4.1	The WT-based Self-optimized Dynamic Routing: What Are the Proper CSI Averaging Percentile and Time Window? . . .	185
7.4.2	Characterization of the FRS Feeder Link Support to the Allo- cated Rate on the WT's Access link for Self-optimized Routing	185
7.4.3	Performance of the PC Mechanism with Subcarrier Packing .	186
7.5	Development of the NRS MAC and PC Techniques in Chapter 5 and Chapter 6 towards femtocells	186
7.6	Terminal-to-terminal Relaying and Ad-hoc Networks	187
A	Generating a Large Number of Independent Time-Frequency Cor- related Fading Realizations in OFDMA Multicellular Networks	189
B	Channel Emulation and Generation of Error Patterns at the WTs and NRSs with lossless NRS-WT links	193
	Bibliography	196

List of Figures

1.1	A partial network in multicellular relay networks.	3
2.1	Static resource allocation in [36] and [40].	21
2.2	Structure and sequence of operation of the integrated radio resource allocation in [42].	23
2.3	Network layout and spatial reuse pattern [48].	30
2.4	Half duplex relaying patterns for SSA and SRA as in [50].	34
2.5	Example relay-enhanced cellular architecture. A central controller performs global resource allocation in centralized conventional networks.	36
2.6	Snap shots of the cellular network in downlink explaining the CCI uncertainty problem: (a) CCI during frame i , (b) CCI during frame $i + 1$. Solid arrows represent desired signals in the center cell.	45
2.7	The behavior of Jain's index function in a 2-users case.	55
2.8	The behavior of the IEEE 802.16m index in a 2-users case.	56
2.9	Time-frequency correlated Rayleigh fading realization for a speed of $v = 10$ Km/hr $f_c = 2.5$ GHz.	59
2.10	Time-frequency correlated Rician (K-factor of 10 dB) fading realization 144 subchannels $f_d = 4$ Hz.	60

3.1	An example multihop packet radio network with multicommodity flows (shown in red, green and blue along with their respective sinks). The slotted time axis shows the traditional quasi-FDR protocol in throughput-optimal scheduling at the nodes N_2 , N_5 , and N_7	65
3.2	Example partial network showing the potential links of the BS and RS_2 on subchannel n	70
3.3	The demand matrix during one iteration. Rows with assigned entries are crossed out and eliminated. Red entries reflect on the queue updates due to the previous iteration.	77
3.4	User and frequency partitioning in the reference relay-aided PFS scheme.	80
3.5	Time-average user throughput as function of user location and shadowing with 25 UTs/cell using 3 and 6 RSs.	83
3.6	CDF of the time-average user throughput with 25 users and 6 RSs per cell.	85
3.7	Lower tail behavior of the CDF of the time-average user throughput with 25 users and 6 RSs per cell.	86
3.8	Total average cell throughput for the competent schemes.	87
3.9	Time-average fairness using the IEEE 802.16m index with different numbers of users and RSs.	88
3.10	Long-term fairness using Jain's index with different numbers of users and RSs.	89
3.11	Normalized histogram of the number of subchannels assigned to transmitting nodes.	91
3.12	CDF of time-average user throughput with 15 users/cell and 6 RSs showing no performance degradation under 50%-less partial feedback.	93
3.13	CDF of the IEEE 802.16m fairness index with 15 users/cell and 6 RSs showing no performance degradation under 50%-less partial feedback.	93

4.1	A representative cell in the multicellular network with asymmetric traffic flows and queue dynamics. The blue and red shades distinguish UTs pertaining to different classes along with their respective queues. . . .	98
4.2	Generic frame structures for; (a) the proposed Variant-A; (b) the investigated Variant-B; and (c) the quasi-FDR in Chapter 3.	99
4.3	Time-average throughput comparison of Variant-A with the reference schemes at 30 UTs/cell.	116
4.4	The CDF of time-averaged user throughput with 40 UTs per cell and emphasis on the lower tail behavior.	117
4.5	The total cell average throughput vs. the number of UTs per cell for the proposed and reference schemes.	118
4.6	Time-average throughput fairness for the proposed scheme and the reference quasi-FDR scheme with 40 UTs per cell.	118
4.7	Long-term throughput fairness for the proposed scheme and the reference quasi-FDR scheme with 40 UTs per cell and a time window of 20 frames.	119
4.8	Time-average user throughput as function of user location and shadowing with 20 UTs/cell with asymmetric traffic and 3 or 6 RSs. Other scatters are not shown for figure clarity.	120
4.9	The CDF of time-averaged user throughput with symmetric and asymmetric traffic.	121
4.10	The time-average absolute and relative fairness with symmetric and asymmetric traffic.	123
4.11	Long-term absolute and relative fairness with symmetric and asymmetric traffic.	123

5.1	The region of interest may represent a cell of the hexagonal grid, or a sector of a cell, in the multicellular network. All resources are available in this region.	131
5.2	Frame structure of the proposed schemes showing 3 active FRSs and the overlaying of the NRS operation from a downlink frame to the following one after an uplink frame has elapsed.	133
5.3	Examples of subchannel pairing between the feeder link of FRS _{<i>m</i>} and the access links of its connected WTs.	135
5.4	The WT-based self-organizing routing or link selection strategy.	139
5.5	Illustration of the cooperation of the serving FRS and the dedicated NRS to assist the WT. <i>S</i> is the number of OFDM subcarriers per subchannel.	141
5.6	Illustration of the operation of the heuristic semi-centralized RRA scheme.	146
5.7	Illustration of the operation of the distributed RRA scheme and the achievable rates matrix at the BS.	150
5.8	CDF of the time-averaged user throughput for the distributed and semi-centralized schemes, using 3 or 6 FRSs, with and without NRSs, $K = K_{nom} = 25$	154
5.9	Cell-edge performance demonstrated by the lower-tail behavior of the CDFs of the time-averaged user throughput for the semi-centralized scheme with and without NRSs, $K = 15$	155
5.10	Scatter plot of the time-averaged user throughput for the distributed and semi-centralized schemes with and without NRSs, $K = 25$	156
5.11	Normalized histogram of the subchannel reuse factor for the distributed scheme.	158

5.12	Normalized histogram of the subchannel reuse factor for the semi-centralized scheme.	159
6.1	NRS power-subchannel allocation for the case where the least possible number of subchannels is used, given the instant of the NRS medium access.	165
6.2	NRS power-subchannel allocation for the case where no packing is employed and potentially more subchannels are used.	166
6.3	Flow chart of the NRS joint power and subchannel allocation algorithm.	169
6.4	CDF plots of the total power saving gain achieved by the RRM scheme with PC for 15 users/cell.	171
6.5	Normalized histograms of the number of subchannels assigned to users during the first sub-frame of the DL frame.	173
6.6	CDF plots of the power saving gain per subchannel achieved by the RRM scheme with PC for 15 users/cell.	174
6.7	Scatter plots of user time-average throughput for the schemes with 15 WTs and 3 FRs. Curve fittings represent the distance-based conditional means.	175
6.8	CDF plots of the instantaneous throughput of directly connected WTs for the PC and NoPC schemes at 6 FRs/cell.	176
A.1	A single run of the time-frequency correlated fading in a simulation drop showing the pointers' positions at time frame t.	190
A.2	An example network layout in a simulation drop.	191

B.1	Reproduced: Illustration of the cooperation of the serving FRS and the dedicated NRS to assist the WT. S is the number of OFDM subcarriers per subchannel. Red arrows indicate the subcarriers received in error at the WT and the NRS.	194
B.2	Matlab code for error pattern generation (channel emulation) and throughput calculation at the receiving WT and NRS assuming lossless NRS-WT links.	195

List of Tables

3.1	Simulation parameters	82
5.1	System parameters	152
5.2	Outage probability based on users' instantaneous throughput rates, $R_{min} = 250$ Kbps.	154
5.3	Mean network reuse factor for the proposed distributed and semi- centralized schemes.	158

Nomenclature

ACK	Acknowledgement
AF	Amplify-and-Forward
AP	Access Point
APC	Adaptive Power Control
AM	Adaptive Modulation
AMC	Adaptive Modulation and Coding
AWGN	Additive White Gaussian Noise
BER	Bit Error Rate
BICM	Bit Interleaved Coded Modulation
B-LDPC	Block Type-Low Density Parity Codes
BLER	Block Error Rate
BPSK	Binary Phase Shift Keying
BS	Base Station
CAC	Connection Admission Control
CAPEX	Capital Expenditure
CBR	Constant Bit Rate
CCI	Co-channel Interference
CDMA	Code Division Multiple Access
CDMA EV-DO	Code Division Multiple Access EVolution Data Optimised
CDF	Cumulative Density Function

CR-QAM	Continuous Rate-Quadrature Amplitude Modulation
CSI	Channel State Information
DF	Decode-and-Forward
DFT	Discrete Fourier Transform
DL	Downlink
DPA	Dynamic Power Allocation
EPA	Equal Power Allocation
FD	Full Duplex
FDD	Frequency Division Duplexing
FDMA	Frequency Division Multiple Access
FDR	Full Duplex Relaying
FEC	Forward Error Correction
FFT	Fast Fourier Transform
FRN	Fixed Relay Node
FRS	Fixed Relay Station
HDR	Half Duplex Relaying
ICI	Inter-Cell Interference
IEEE	Institute of Electrical and Electronics Engineers
ILP	Integer Linear Programming
IP	Integer Programming
KKT	Karush-Kuhn-Tucker conditions
LU	Lower Upper (factorization)
LOS	Line-of-Sight
LP	Linear Programming
MAC	Medium Access Control
MC-CDMA	Multicarrier-Code Division Multiple Access
MH	Multihop

MIMO	Multiple Input Multiple Output
MMPP	Markov Modulated Poisson Process
MQAM	M-ary Quadrature Amplitude Modulation
MRC	Maximal Ratio Combining
MS	Mobile Station
NGN	Next-Generation Network
NLOS	Non-Line-of-Sight
NLP	Non-Linear Programming
NRS	Nomadic Relay Station
NSM	Network Simplex Method
OFDM	Orthogonal Frequency Division Multiplexing
OFDMA	Orthogonal Frequency Division Multiple Access
OPEX	Operational Expenditure
PC	Power Control
PDF	Probability Density Function
PN	Pseudo Noise
QoS	Quality of Service
RAM	Random Access Memory
RAN	Radio Access Network
RNC	Radio Network Controller
RRA	Radio Resource Allocation
RRM	Radio Resource Management
RS	Relay Station
SC-CDMA	Single Carrier-Code Division Multiple Access
SC-FDMA	Single Carrier-Frequency Division Multiple Access
SC-TDMA	Single Carrier-Time Division Multiple Access
SDMA	Spatial Division Multiple Access

SH	Single-hop
SIR	Signal-to-Interference Ratio
SNR	Signal-to-Noise Ratio
SINR	Signal-to-Interference-plus-Noise Ratio
SON	Self-Organizing Network
TDD	Time Division Duplexing
TDM	Time Division Multiplexing
TDMA	Time Division Multiple Access
TGPS	Truncated Generalized Processor Sharing
UL	Uplink
UMTS	Universal Mobile Telecommunications System
UT	User Terminal
VBR	Variable Bit Rate
WF	Water Filling
WG	Work Group
WS	Wireless Station
WT	Wireless Terminal
W-CDMA	Wideband Code Division Multiple Access
WINNER	Wireless World Initiative New Radio

Chapter 1

Introduction

Next-generation wireless communication networks are expected to provide ubiquitous high-data-rate coverage in the most cost-effective manner. To achieve this objective, high-spectral-efficiency schemes are required in conjunction with aggressive resource reuse strategies to ensure prudent use of scarce radio resources. Interest in orthogonal frequency-division multiplexing (OFDM) is growing steadily, as it appears to be a promising air-interface for the next generation of wireless systems due, primarily, to its inherent resistance to frequency-selective multi-path fading and the flexibility it offers in radio resource allocations. Exploiting the dynamism in wireless channels, OFDM subcarriers can be adaptively modulated and/or assigned to the “best” link to achieve frequency and multi-user diversity gains. OFDMA results from using OFDM in conjunction with frequency-division multiple access protocol, in which a user may be assigned one or more subcarriers in order to meet its communication requirements.

Wireless coverage is only ubiquitous if reliable service is provided throughout the serving area of a base station (BS). It is advantageous for network service providers to distribute system capacity across the network area, reaching users in the most cost-effective way. With the traditional cellular architecture, increasing the capacity along with the coverage requires deployment of a large number of BSs. This approach is cost prohibitive to network service providers [1], [2]. As an alternative, relaying

techniques are expected to alleviate this coverage problem since the relay station (RS) with less functionality than the BS can forward high data rates to remote areas of the cell while lowering infrastructure cost. The synergy between OFDMA and relaying techniques offers a promising technology for providing high data rate to users everywhere, anytime.

To this end, an OFDMA-based relay-enhanced network comprising various forms of infrastructure-based or dedicated relays is envisaged in the next-generation networks. These relays could be owned, leased or given at no cost to users by the service provider. How to perform radio resource management (RRM) in such a complex environment is not immediately clear. However, there is a consensus on the need to devise intelligent RRM algorithms with excellent bandwidth efficiency that can harness the full potentials of relay-enhanced OFDMA-based networks. The literature reveals that RRM algorithms have been extensively investigated for conventional OFDMA-based networks (without relays), e.g., [3] - [7]: it is only recently that attention has started to shift to relay-based OFDMA multi-cell networks.

1.1 The RRM Problem Statement

RRM algorithms are usually designed to exploit the variations in wireless channels by adaptively distributing scarce communication resources to either maximize or minimize some network performance metrics. In relay-enhanced OFDMA-based wireless networks, a typical resource allocation problem statement might be as follows:

“How many information bits, how much transmit power (for users, relays, or BS), and which subcarriers (or subchannels) should be assigned to different links to either maximize or minimize a desired performance metric, e.g., system throughput (capacity) or total transmit power in the network, respectively?”

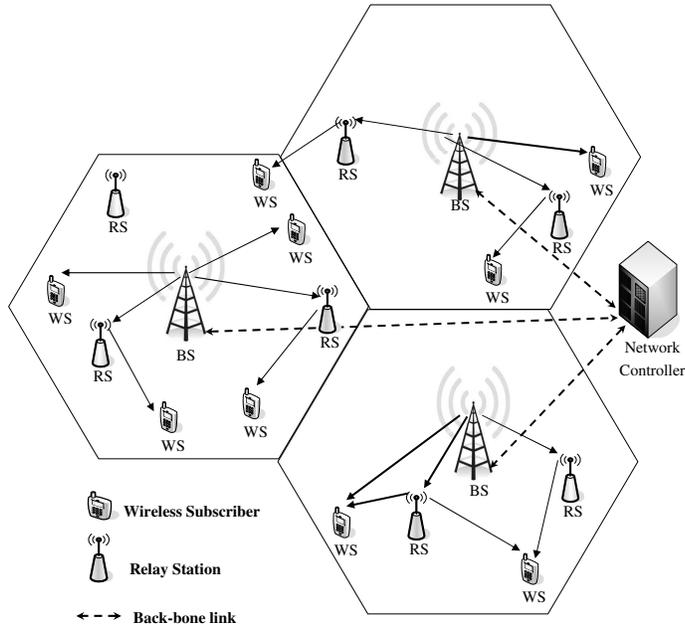


Figure 1.1: A partial network in multicellular relay networks.

In a situation where a network device has “free” access to energy (e.g., fixed relay applications, where off-the-wall power is always available), the main objective of the RRM algorithm in downlink transmissions could simply be to maximize the system throughput. To this end, scheduling, routing, bit-loading and adaptive modulation constitute some of the tools that are commonly employed in RRM solutions.

An Optimization Example [8]: Let us consider formulating the downlink sum-rate-maximizing subcarrier allocation problem for a half-duplex relay network, assuming that the average power (P) for each subcarrier is the same and fixed. Consider a single cell from the network shown Fig. 1.1 where the wireless station (WS) and RS are treated as contending for radio resources, an optimization problem can be expressed in the following way. Let M , K , and N represent the number of relays, WSs, and subchannels, respectively. Let $\rho_{m,n}$ represent subcarrier n assignment variable, indicating whether subcarrier n is allocated to the BS - RS_m link ($m = 1, \dots, M$) or BS - WS_m link ($m = M + 1, \dots, M + K$). Let $\rho_{m,k,n}^*$ also indicate that subcarrier n is allocated to the BS ($m = 0$) or RS_m ($m = 1, \dots, M$) - WS_k link during the

second time slot of the downlink frame. Let $\gamma_{m,n}$ and $\gamma_{m,k,n}^*$ denote the signal-to-interference-plus-noise ratio (SINR) in the corresponding transmissions. Assuming that the channel does not change within the duration of the two-hop transmission, we can express $\gamma_{M+j,n} = \gamma_{0,j,n}^*$ ($j = 1, \dots, K$), which will lead to the subcarrier allocation problem formulated in [8] as:

$$\max_{\rho, \rho^*} \frac{1}{2} \sum_{m=1}^M \min \left\{ A_m, B_m \right\} + \frac{1}{2} \left(\sum_{m=M+1}^{M+K} A_m + B_0 \right), \quad (1.1)$$

subject to:

$$\begin{aligned} \sum_{m=1}^{M+K} \rho_{m,n} &= 1, \quad \forall n, \\ \sum_{m=0}^M \sum_{k=1}^{M+K} \rho_{m,k,n}^* &= 1, \quad \forall n, \\ \rho_{m,n} &\in \{0, 1\}, \quad \forall m, n, \\ \rho_{m,k,n}^* &\in \{0, 1\}, \quad \forall m, k, n, \end{aligned}$$

where $A_m = \sum_{n=1}^N \rho_{m,n} \log_2(1 + \gamma_{m,n}P)$ and $B_m = \sum_{n=1}^N \sum_{k=1}^K \rho_{m,k,n}^* \log_2(1 + \gamma_{k,m,n}^*P)$.

The first term in the cost function (1.1) represents the achievable rate of data flow from BS to WSs via RSs and the second term stands for the achievable rate in the direct channels from BS to WSs.

One main challenge confronting some formulations of RRM algorithms as presented in [8] and [9], is that they fall in the class of integer programming problems that incur a prohibitive search complexity of $\mathcal{O}((MK)^N)$ [9]. The complexity challenge is primarily due to combinatorial nature of the OFDMA and the optimization constraints. The constraints above can be interpreted as a subcarrier can only be assigned to one transmitter-receiver pair, meaning that there is no sharing or intra-cell subcarrier reuse.

Radio resource management algorithms are sometimes classified according to their output objectives or processing requirements. In terms of their output, the algorithm can be classified as *network-centric*, achieving high capacity from the perspective of the service provider, or *user-centric*, providing reliable service and fair share of resources to the user, or having flavors of both. On the other hand, considering the processing requirements, the classifications are 1) *centralized*, where the algorithm requires, in principle, global or centralized knowledge of all interference and channel state information (CSI) for all nodes (users, relays) in the network, or 2) *semi-distributed*, where the algorithm requires a limited global knowledge of network link conditions or 3) *fully distributed* where the algorithm provides each resource-allocating node the ability to allocate resources based on local channel conditions. While the centralized schemes constitute the bulk of RRM literature as perceived optimal in principle, the high signaling overhead (which consumes part of the system throughput) and complexity incurred in such proposals make them less attractive for OFDMA-based relay networks; therefore, distributed RRM algorithms are expected to attract more attention for these type of networks. Nevertheless, an interesting question requires an answer; can we design centralized RRM algorithms that incur low complexity and save substantially in feedback overhead?

We attempt to answer this question in Chapters 3 and 4 whereas distributed and semi-centralized self-organizing schemes are devised in Chapters 5 and 6.

1.2 Publications, Patent filings, and Technical Reports

The work that has been published, accepted, or submitted is presented in Chapters 2 - 6 whereas Chapter 7 summarizes the thesis conclusions and contributions and discusses the possible extensions as well. Several articles of archival value have been produced and they are cross-referenced to individual chapters as follows:

- Chapter 2 serves as a literature survey and tutorial. The contents of that chapter are published in following journal articles and technical report:
 - **M. Salem**, A. Adinoyi, H. Yanikomeroglu, and D. Falconer, “Opportunities and challenges in OFDMA-based cellular relay networks: A radio resource management perspective,” *IEEE Transactions on Vehicular Technology*, 59(5), pp. 2496-2510, January 2010.
 - **M. Salem**, A. Adinoyi, M. Rahman, H. Yanikomeroglu, D. Falconer, Y.-D. Kim, E. Kim, and Y.-C. Cheong, “An overview of radio resource management in relay-enhanced OFDMA-based networks,” *IEEE Communications Surveys and Tutorials*, 12(3), pp. 422-438, Third Quarter 2010.
 - **M. Salem**, A. Adinoyi, M. Rahman, H. Yanikomeroglu, and D. Falconer, Radio Resource Management in OFDMA-based Fixed Relay Networks: Literature Review. Deliverable Y1-D1, submitted to Samsung Electronics, Korea, (49 pages), 15 December 2007.
- Chapter 3 presents our first fairness-aware RRM scheme designed for fixed relay networks using the traditional relaying protocol in the relevant literature. The contents of that chapter are published in following journal paper, patent fillings, conference paper, and technical report:
 - **M. Salem**, A. Adinoyi, M. Rahman, H. Yanikomeroglu, D. Falconer, and Y.-D. Kim, “Fairness-aware radio resource management in downlink OFDMA cellular relay networks,” *IEEE Transaction on Wireless Communications*, 9(5), pp. 1628-1639, May 2010.
 - **M. Salem**, A. Adinoyi, M. Rahman, H. Yanikomeroglu, D. Falconer, and Y.-D. Kim, Apparatus and Method for Allocating Subchannels and Controlling Interference in OFDMA Systems. Filed by Samsung, Korea patent application no: P2008-0054726 (filing date: 11 June 2008), US patent application no: 12/341,933 (filing date: 22 December 2008), international patent application no:

PCT/KR2009/002119 (filing date: 23 April 2009).

- **M. Salem**, A. Adinoyi, M. Rahman, H. Yanikomeroglu, D. Falconer, Y-D. Kim, W. Shin, and E. Kim, “Fairness-aware joint routing and scheduling in OFDMA-based cellular fixed relay networks,” *IEEE International Conference on Communications (ICC)*, June 2009.
- **M. Salem**, A. Adinoyi, M. Rahman, H. Yanikomeroglu, and D. Falconer, Radio Resource Management through Joint Routing and Fair Scheduling and Multi-cell Coordination in OFDMA Fixed Relay Networks. Deliverable Y1-D2, submitted to Samsung Electronics, Korea, (50 pages), 15 March 2008.
- In Chapter 4, a different and more practical design improving the performance of the first scheme under higher traffic loads is presented. The contents of that chapter are presented in the following journal paper, patent fillings, and conference paper:
 - **M. Salem**, A. Adinoyi, H. Yanikomeroglu, and Y.-D. Kim, “Fair resource allocation towards ubiquitous coverage in OFDMA-based cellular relay networks with asymmetric traffic,” accepted to *IEEE Transactions on Vehicular Technology*, January 2011.
 - **M. Salem**, A. Adinoyi, H. Yanikomeroglu, D. Falconer, and Y.-D. Kim, Method for Performing Fair Resource Allocation in OFDMA-based Relay-Networks. Filed by Samsung, Korea, patent application no: P2009-0022132 (filing date: 16 March 2009), USA patent application no: 12/567,776 (September 2009), and international application is underway.
 - **M. Salem**, A. Adinoyi, H. Yanikomeroglu, D. Falconer, and Y.-D. Kim, “A fair radio resource allocation scheme for ubiquitous high-data-rate coverage in OFDMA-based cellular relay networks,” *IEEE Global Communications Conference (Globecom)*, December 2009.
- A decentralized RRM scheme for networks enhanced by both fixed and nomadic

relays is presented in Chapter 5. That work is presented in the following journal paper, invention disclosure, conference paper, and technical reports:

- **M. Salem**, A. Adinoyi, H. Yanikomeroglu, “Integrating self-organizing nomadic relays into OFDMA fixed-relay cellular networks”, submitted to *IEEE Transactions on Mobile Computing*, no. TMC-2010-10-0487, 19 October 2010.
 - **M. Salem**, A. Adinoyi, H. Yanikomeroglu, and Y.-D. Kim, Method for Relaying Data in Wireless Network and Personal Relay of Enabling the Method, and Mobile Device for Communicating with the Personal Relay. Filed by Samsung, Korea; Korea patent application no: P2009-0084026 (application date: 07 September 2009).
 - **M. Salem**, A. Adinoyi, H. Yanikomeroglu, and Y.-D. Kim, “Radio resource management in OFDMA-based cellular networks enhanced with fixed and nomadic relays”, *IEEE Wireless Communications and Networking Conference (WCNC)*, April 2010.
 - **M. Salem**, A. Adinoyi, H. Yanikomeroglu, and D. Falconer, Distributed Radio Resource Management in OFDMA-based Multi-cellular networks Enhanced with Fixed and Nomadic Relays. Deliverable Y2-D1, submitted to Samsung Electronics, Korea, (59 pages), 15 January 2009.
 - **M. Salem**, A. Adinoyi, H. Yanikomeroglu, and D. Falconer, Distributed Radio Resource Management in OFDMA-based Multi-cellular Networks Enhanced with Fixed and Nomadic Relays. Deliverable Y1-D3, submitted to Samsung Electronics, Korea, (32 pages), 15 September 2008.
- Chapter 6 presents a novel joint power and subchannel allocation algorithm for the nomadic relays of the decentralized RRM schemes. That work is presented in the following conference paper and technical report:
 - **M. Salem**, A. Adinoyi, H. Yanikomeroglu, and Y.-D. Kim, “Nomadic relay-directed joint power and subchannel allocation in OFDMA-based cellular fixed

relay networks”, *IEEE Vehicular Technology Conference (VTC-Spring)*, May 2010.

- **M. Salem**, A. Adinoyi, H. Yanikomeroglu, and D. Falconer, Distributed radio resource management with power control in OFDMA-based multi-cellular networks enhanced with fixed and nomadic relays. Deliverable Y2-D2, submitted to Samsung Electronics, Korea, (34 pages), 15 April 2009.

A journal paper with the following interim title is in preparation:

- **M. Salem**, A. Adinoyi, H. Yanikomeroglu, and Y.-D. Kim, “Joint power and subchannel allocation for the self-organizing nomadic relays in OFDMA-based cellular fixed-relay networks,” to be submitted to an IEEE journal, January 2011.

Chapter 2

Overview on RRM in OFDMA Relay Networks

This chapter provides a survey of the current literature on OFDMA networks enhanced with decode-and-forward relaying and provides their link to earlier literature in non-OFDMA networks. We address some of the opportunities, challenges, and terminologies associated with the migration from conventional cellular architecture to relay-enhanced in OFDMA-based networks. We highlight the fairness concerns in such networks and discuss some fairness metrics as well as possible fairness implementations in radio resource allocation (RRA) algorithms. The work presented in this chapter has appeared in the journal articles [10] and [11] and the technical report [12].

The RRM problem example discussed in Section 1.1 and the discussions in Chapters 3 and 4 are based on a similar topology. In systems similar to that depicted in Fig. 1.1, the BS continuously keeps track of the quality of some or all links (e.g., SINR per subcarrier). For slowly varying channels, BSs or RSs can be assumed to have sufficiently accurate estimate of the channel states. The base stations can optimize the resource allocation locally or they can involve a central or network controller as shown in Fig. 1.1. The radio resources are assigned to links based on a number of considerations with the overall aim of improving the throughput. User fairness can also be an important aspect of system performance. The BS directly or through RSs serves the WSs in each cell. The data of each WS in a cell can potentially be routed

through any of the relays according to the routing criterion. All cells share the same spectrum if no clustering or fractional reuse is employed. It is also observed that cooperation (where a WS is served through more than one link) can be invoked to deliver a reliable service to WSs.

2.1 Activities on Relay-based Architecture among Standardization Bodies, Forums, and Consortiums

To the author's knowledge, the first standard involving relays in the form of analog repeaters was introduced in personal wireless terminal (PWT) interoperability standard in the 90's. PWT emerged from ETSI digital enhanced cordless telecommunications (DECT) standard as a North American TIA variant. Operated in the 1900 MHz band using time-division duplex (TDD) and time-division multiple access (TDMA), these relays were to extend range up to a few kilometers for voice and low-rate data services [13]. In recent years, as one of the key technologies for 4G networks, relaying has again attracted tremendous attention in order to obtain coverage/capacity enhancement as well as to extract benefits such as cooperative diversity gains.

The idea of relaying has previously been examined in the 3GPP in the form of ad-hoc peer-to-peer multihop protocol known as opportunity-driven multiple access (ODMA). In the ODMA protocol, the cell-edge wireless stations (WSs) communicate with the BS through the help of other WSs that are closer to the BS. However, the ODMA protocol did not make it to the standard due to complexity and signalling overhead concerns [14]. Nevertheless, a number of lessons were learned. Currently, numerous standardization bodies, forums, and consortiums have been working on relay-based architectures in scenarios ranging from wireless local area networks (WLAN) to wide-area cellular networks.

Early WLAN standards, such as IEEE 802.11a/b/g, do not support relay-based deployment due to their contention-based multiple access technique. Researchers

proposed a number of possible relay-friendly modified MAC protocols, for instance, relay-enabled point coordination function (rPCF) [15] and relay-enabled distributed coordination function (rDCF) [16], among others, to enable the use of relays with these standards. However, Task Group “S” (TGs) of the IEEE 802.11 Work Group (WG), formed in 2003, released an unapproved draft version of IEEE 802.11s wireless mesh network (WMN) in March 2007, which is an amendment to support mesh and relay based WLAN. The key element in this WLAN version is called mesh point (MP). Being very similar to the traditional access point (AP), MP can connect mesh APs, other MPs, and STations (STA) wirelessly. An excellent overview on 802.11s is available in [17]. European ETSI HIPERLAN/2, on the other hand, has shown an ability to support relay-based deployment due to its dynamic TDMA access on an OFDM air-interface.

IEEE 802.16m and 3GPP advanced long term evolution (LTE-A), which contend beyond 3G standards are the two most notable standards to advocate relaying. Although IEEE 802.16m inherits multihop relaying from predecessor IEEE 802.16j standard, the evolving LTE-A is expected to incorporate multihop relaying to further narrow down the gaps between these two standards. The worldwide interoperability for microwave access (WiMAX) forum was created in 2001 to promote conformance and interoperability of products based on harmonized IEEE 802.16/ETSI HiperMAN standard. Like IEEE WLAN, the focus of IEEE 802.16 early standards was not on relaying until IEEE 802.16e; in fact, this is the first standard that supports mobility along with basic mesh networking capability. IEEE 802.16j mobile multihop relay is an amendment to IEEE 802.16e standard focusing on coverage/capacity expansion. The first unapproved draft version [18] has been published by the end of 2008. In IEEE 802.16m, OFDMA in TDD has been used for both the downlink and uplink while OFDMA and single carrier frequency-division multiple access (SC-FDMA) are the access schemes for the downlink and uplink of LTE, respectively. Adaptive

OFDMA is proposed for the downlinks in 3GPP-LTE or Evolved UTRA. To this end, an excellent overview on design issues relating to adaptive transmission for OFDMA-based systems beyond 3G can be found in [19].

In addition to these standardization efforts, European 6th framework projects such as FIREWORKS¹ and WINNER² have been involved in the 4G (and beyond) wireless research. They consider relaying and OFDMA as essential elements required to meet the demand of future wireless services. Accordingly, RRM design occupies a pivotal position in these projects. We now briefly discuss some of the basic tools employed in designing RRM solutions.

2.2 Some Basic RRM Tools

2.2.1 Scheduling

Different scheduling policies are available for the selection of users to be served based on different network criteria. The most common scheduling algorithms are the following: 1) *Round-robin scheduler*: This user-centric and fairness-conscious scheduler assigns the same amount of physical resources to users in turn. Conventional round-robin scheduler, probably the most natural form of scheduling, does not guarantee quality of service since, it neither utilizes the queue state nor exploits the channel variability in the scheduling policy, thereby sacrificing the inherent multiuser diversity and achievable network capacity. 2) *Max-SINR scheduler*: This network-centric scheduler is the best in terms of total capacity maximization at the price of fairness, as it fully exploits multiuser diversity inherent in the network. 3) *Proportional-fair scheduler*: This scheduler provides an intermediate solution that realizes the multiuser diversity gains while maintaining fairness across users. At any transmit node, this scheduler allocates a subchannel to the WS that maximizes the ratio of its achievable

¹<http://www.ist-fireworks.eu>

²www.winner-ist.org

rate on that subchannel to its exponentially weighted average rate. It is observed in [20] that proportional fair scheduling does not guarantee queue stability even for low traffic loads.

The fair throughput and early-deadline-first schedulers are also discussed in many publications in the literature, including [21], [22], and [23]. The early-deadline-first scheduler is particularly suitable for real-time applications since it can support or allow priority service. A survey of the state-of-the-art adaptive RRA and scheduling approaches in conventional OFDMA-based systems is provided in [24].

Unless modified, the scheduling techniques mentioned above may not deliver optimum performance in the future OFDMA-based relay networks that have many optimization parameters [25]. This fact is recognized in one variant of the round-robin scheduler known as the exhaustive round-robin scheduler, where bursts of un-equal resources are assigned to users. The scheduler continues to serve a user until its queue is exhausted. In this way the overhead requirement is reduced while fairness is sacrificed for increased network capacity.

2.2.2 Routing

In relay enhanced wireless networks, scheduling and routing resource management decisions should not be isolated. Routing can be viewed as the process of establishing efficient connectivity between nodes over multihop links. In communication systems with large number of nodes, there will be diverse link qualities associated with the various links along any route. Therefore, different routing schemes are expected to produce different performance results and overhead [26]. Performing routing and scheduling jointly is known to produce superior performance results compared to decoupled scheduling and routing. A review of some routing metrics and algorithms is available in [27].

2.2.3 Link Adaptation: Adaptive Modulation and Coding

The best modulation level to employ for transmission can be obtained by adapting modulation according to the instantaneous SNR to achieve the maximum possible transmission rate for a particular target bit error rate (BER); this process is known as adaptive modulation. The combined use of adaptive modulation and channel coding techniques has given rise to what is commonly known as adaptive modulation and coding (AMC). The SNR can be partitioned into a number of consecutive non-overlapping intervals with boundary points obtained by re-arranging the rate expression [28]

$$r_{i,j,n} = W \log_2 \left(1 + \frac{-1.5 \gamma_{i,j,n}}{\ln(5 \text{ BER})} \right), \quad (2.1)$$

where $r_{i,j,n}$ is the achievable rate, $\gamma_{i,j,n}$ is the received SINR from source i at destination j on subchannel n and W is the OFDM subchannel bandwidth. Evaluating the number of bits (spectral efficiency) to employ for the OFDM symbols of a subcarrier (or subchannel) is sometimes referred to as *bit loading*.

2.2.4 Power Control

Transmit power control (PC) is an important interference combatting mechanism thereby constituting a means for improving the network throughput by reducing the co-channel interference and facilitating frequency reuse. The RRM schemes earmarked for the future wireless systems are envisioned to be aggressive in reusing the scarce licensed spectrum enabling the operation of many active devices ranging from nomadic relays [18], [29] to femtocell access points [30]. Although dynamic power control (DPC) has been a key player in resource assignment in CDMA-based systems, it has been concluded that centralized power control is not practical in multicarrier/OFDMA systems comprising many active nodes as it requires accurate measurements of all gains in all radio links. Rather, practical power control schemes

have to rely on far less accurate measurements and allow only a limited information to enable distributed operation. In such case, DPC could be also employed to ensure that a wireless transmitter uses just enough power to facilitate its operation without sacrificing system performance or causing receiver saturation. Therefore, PC will continue to be a relevant tool. Even systems based on well-designed RRM schemes could still benefit from PC techniques in addition to their traditional function as interference combatting mechanisms.

2.2.5 Cooperative Relaying in OFDMA Networks

The benefits of cooperative communication for addressing physical layer problems have been well investigated. It is only recently that the cross-layer interaction of cooperative schemes has started to receive attention for the emerging OFDMA-based networks. In fact, cross-layer optimization is identified as an important strategy to ensure overall system performance in wireless networks [31]. The authors in [32] and [33] introduce cross-layer optimization to the resource allocation in OFDMA-based relay networks by incorporating a physical layer cooperative technique into the RRM problem formulation. Their treatments are similar in many respects, except that [33] considers multihop scenario (more than two hops) while [32] treats only a two-hop case. Furthermore, [33] considers multi-modal relaying (system selects between amplify-and-forward or decode-and-forward relaying, depending on the channel conditions and power allocation), while [32] uses only decode-and-forward relaying. While [32] maximizes the minimum throughput among all BSs under routing and PHY constraints, [33] maximizes network sum utility by optimally choosing the active data stream and allocating power in each tone in conjunction with selecting the best relay node and the best relaying method.

The schemes described in [32] and [33] need to operate in a slow-fading environment to be able to acquire the CSI for implementing the adaptive power and

bit-loading in a centralized manner. For this reason, the receiver needs to reliably estimate channel and convey it to the BS. The authors also assume that the OFDM frames are synchronized throughout the network to enable nodes' cooperation at the level of OFDM subcarrier [33]. In an effort to reduce the computational complexity, some conditions are imposed such as the source-relay, source-destination, and relay-destination links all use the same subcarrier. This restriction results in sacrificing part of the inherent frequency diversity in OFDMA systems. Furthermore, the challenges in ensuring such a network-wide synchronization have not been addressed, and the impact on the proposed system is not discussed. Given that OFDM/OFDMA techniques are very sensitive to frequency and timing synchronization errors [34], such an assumption might be too stringent a condition. The impact of these errors on the system performance cannot be left to speculation. Therefore, we believe that further investigation of these issues is required. The performance degradation could result from the loss of orthogonality among subcarriers leading to interference (caused by the inaccurate compensation of frequency offset) or inter-symbol interference due to timing error. Thus, a sufficiently long cyclic prefix or guard interval between adjacent OFDMA symbols may be required to provide protection against these errors. This additional overhead has the effect of reducing the achievable network throughput.

To further emphasize the need for cooperative protocols in the future wireless standards, the IEEE 802.16m working group is considering protocols to enable cooperative communication that is cross-layer optimized. For example, [35] proposes a functional block called cooperative transmission management unit on the upper medium access control. Their description stresses the interaction between the cooperative transmission management unit, scheduling and resource allocator, and the physical layer control, such as interference management, ranging, and link adaptation.

2.3 Earlier RRM Research in Relay Networks

The need to efficiently utilize the scarce radio spectrum has driven the research trend from fixed reuse patterns to dynamic and very aggressive resource allocation. Even in a non-OFDMA air-interface, the existence of relays presents some challenges in RRM design. Therefore, in the following section we present some of the previous RRM schemes that constitute the building blocks of the discussions in the rest of this survey. We start with basic, fixed resource allocation schemes in multicellular networks.

2.3.1 Relaying in Downlink Multicellular Networks

To a large extent, static resource allocation is the common practice in the literature. In particular [36] investigates a static resource allocation in a non-CDMA multi-cell network with 6 fixed relays in each cell. In conjunction with this static allocation, a “pre-configured” relaying channel selection algorithm is used to reduce interference. This algorithm has built-in channel reuse. The reuse is performed in a controlled manner to prevent co-channel interference from increasing to unacceptable levels. This means that a relay is not allowed to reuse any channel in the same cell, but it could reuse the channel from the cell farthest from it (because a cell farthest from a relay most probably has the least co-channel interference to this relay transmission). Note that no channel is reserved for exclusive use of relays, whenever the need arises for relaying, a relay channel is selected from any adjacent cell in a way that keeps the interference within the acceptable level. Different design and path selection criteria are discussed in [26], [36] and [37] in different flavors and forms. Since no channel (frequency) is assigned to relays a priori in these algorithms, the benefits of relaying are achieved without sacrificing system bandwidth.

A centralized downlink joint scheduling and routing in a single cell CDMA EV-DO (EVolution Data Optimized) system is proposed in [38]. Fairness among users is attained through a queue stability strategy. Relaying is done by reusing the same licensed band used in BS-RS links whereas in some of the earlier works, RS-WS links are assigned an unlicensed band such as for IEEE 802.11. The co-channel interference, a form of intra-cell interference, results from concurrent transmissions of different nodes on the same carrier frequency with different pseudo noise (PN) codes. Relays behave like user terminals in uplink and feedback to the BS the rate-control information, as well as the queue size information for the assigned user for each time slot. The joint routing and scheduling examines each possible set of simultaneous active transmitters in a computationally demanding search. For each set of active transmitters, interference power can be calculated at each receiving destination of each link in the set of all possible links defined on the former active set. Relays are allowed to transmit to either WSs or to any other inactive relays during the time frame. The joint scheduling and routing algorithm [38] employs a demand metric that is calculated for candidate links per each set of active transmitters. This metric is proportional to Shannon capacity and to the maximum difference in queue lengths between source node and destination node across each link. The set of transmitting nodes that achieves the highest demand sum, over all its active links, is the optimal set. Implicitly, the user that has the largest difference in queue lengths over any active link is scheduled on that link.

A reduced-complexity implementation of this algorithm was also introduced in [38]. In this approach, the infeasible combinations of active links per each set of active transmitters are eliminated. The infeasible combinations violate the following constraints:

- A relay cannot transmit and receive simultaneously.
- Any WS cannot receive simultaneously from two or more different nodes.

- Any transmitting node can only transmit to a single node per time frame.

The user mobility with such a routing and scheduling algorithm also suggests that data buffered in a relay could be lost if not forwarded to the new serving relay.

In [39], a fixed relay-enhanced multi-cell downlink scenario in FDD mode with TDMA is considered, where the fixed relays are located at the centre of each sector of the cell. By virtue of their relative locations (neighbourhood), antenna directivities, and gains, the cells are organized into in-group BSs representing cells that are mutually interfering with each other. The transmission time frame is divided between the BSs in the dominant interferer group and RSs. In order to schedule and route packets such that dominant interference is minimized, the interference management scheme proposed in [39] utilizes inter-cell coordination among the in-group BSs taking co-channel interference into account. WSs and RSs in a cell estimate their CSI of the links from the serving BS and the in-group BSs and RSs and forward them to the serving BS. From there, they are exchanged among the in-group BSs for informed scheduling decisions. The information exchange can be viewed as a command. Thus, a particular BS might receive different commands from two different groups. When this happens, the BS takes the more conservative command, with the aim to neither cause too much interference to the transmissions of other in-group transmissions nor receive too much interference from other in-group BSs. The objective is to maximize the total network throughput with the least interference suppression. The proposed scheme finds a combination of in-group BS transmissions in the first sub-frame, in addition to the transmission parameters for RS-WS transmission in the next sub-frame. The interference problem is addressed by frequency reuse partitioning from the outset. This a priori resource partitioning is sometimes used in OFDMA networks as well. The downside of this strategy is that it restricts the opportunity in the channel that can be exploited through OFDMA tones. It also requires frequency planning, which can be time-consuming and expensive.

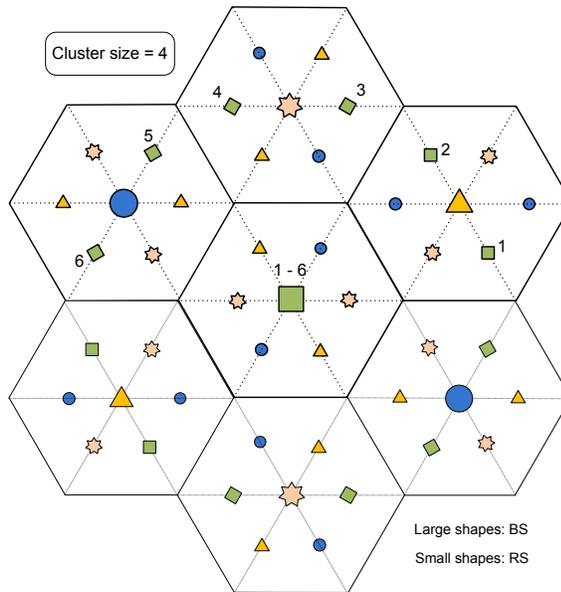


Figure 2.1: Static resource allocation in [36] and [40].

2.3.2 Relaying in Uplink Multicellular Networks

In [40], the static allocation proposed in [36] is extended to FDMA/TDMA in FDD mode for multicellular networks. In contrast to the downlink of [36], an uplink scenario is considered. The objective is to maximize the received SIR values at both BSs and each RS in the uplink. This is achieved by a static allocation that minimizes the interference power received from co-channel terminals. In a cluster size of four cells (the reuse partitioning), the total resources (24 pairs of channels, uplink and downlink) are reused in each cluster; six pairs of channels are assigned to each BS. One pair from the 6 channels of other BS is assigned to each of the 6 RS. Figure 2.1 shows the layout for the channel allocation, where nodes (BS or relay) using the same resources are designated with the same shape. The assumed fixed-power allocation calculates the WS's power level to guarantee the minimum required received power from users located at cell edges. WSs are assumed to be fixed at the worst interfering positions to the cell under investigation. This fixed allocation obviously limits the performance gains, due to the suboptimal frequency reuse partitioning. It also

sacrifices the inherent multiuser diversity and traffic diversity gains.

In a related study to [40], the authors in [41] investigated the performance of an FD/TDMA-based multihop fixed cellular network (in which wireless terminals act as relays whenever necessary) with respect to the number of frequency channels. The outage probability, connectivity, and average node throughput are analyzed to explore the dependency of the performance on the number of frequency channels. Results show that implementing multihop provides not only significant coverage boost but also high node throughput, as long as there are enough frequency channels. It is observed, on the other hand, that if the network has a limited number of frequency channels, no performance gain can be achieved particularly at high loading values.

In [42], a centralized uplink Universal Mobile Telecommunications System (UMTS) cellular network in FDD mode with Wideband-Code Division Multiple Access (W-CDMA) air interface is considered. The scheme relies on integrated radio resource allocation, where routing is coupled with scheduling. In this scheme, WSs use the same carrier frequency to transmit to either the BS or an RS while the RS uses two different carrier frequencies on the two hops. The proposed algorithm is composed of three entities as shown in Fig. 2.2: a routing entity, called *load based route manager*; a resource scheduling entity, called *base station resource scheduler*; and a refinement entity, called *relay station load balancer*, that fine-tunes resources assigned to users. This refinement process is necessary to prevent overloading the RSs. The proposed scheme uses a path-loss-based benchmark relaying scheme in conjunction with a conventional scheduler. Their optimization cost function, based on system consumed-capacity, is different from the distance and path-loss-based cost functions in [37]. Reference [42] introduces a load-cost indicator, defined as the ratio between the consumed system capacity on a particular route (represented by the interference caused by this WS) to its data rate.

Fig. 2.2 shows the three entities that are fed with the path-loss and interference

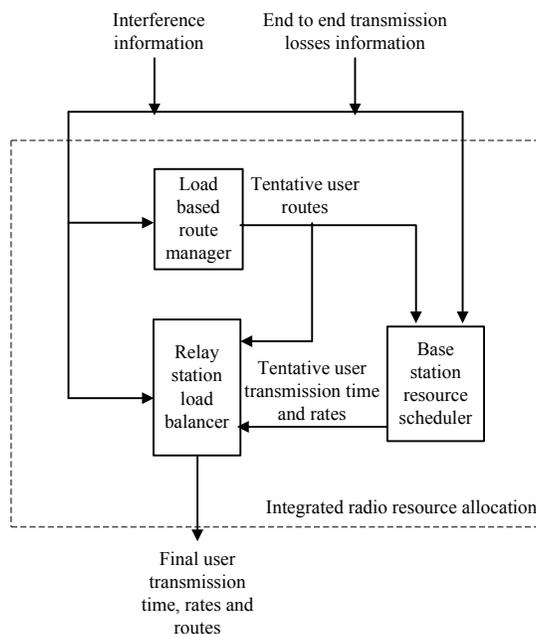


Figure 2.2: Structure and sequence of operation of the integrated radio resource allocation in [42].

information. The load-based route manager prepares tentative user routes based on selection of the routes with least load-cost indicator; then, the base station resource scheduler prepares tentative user transmission times and rates. The outputs of both entities are input to the relay station load balancer that produces the final user rates, times and routes such that no relay is overloaded. The relay station load balancer prioritizes users by the ratio between the cost of direct-link to that of relayed-link. Given that integrating routing to RRM provides superior results, it is observed that integrated radio resource allocation outperforms the reference schemes in several ways. The load-cost indicator reflects the route quality in interference limited scenarios adequately; it prevents overloading of RSs; routing and scheduling are done jointly; and load cost indicator periodically executes its algorithm, thus coping well with the system dynamics.

2.4 RRM in OFDMA Fixed-Relay Networks

There is a noticeable similarity between RRM algorithm design for relay-enhanced OFDMA-based networks and earlier RRM algorithms for non-OFDMA relay networks. Essentially, these algorithms are aimed at ensuring prudent utilization of communication resources, and at improving system performance. The discussions in Section 2.3 (for earlier RRM in non-OFDMA) present the links to the literature in RRM in OFDMA relay networks. The following discussion is grouped into centralized and distributed RRM schemes.

2.4.1 Centralized RRM Schemes in Single-cell OFDMA Relay Networks

A downlink single-cell network with a single fixed RS is considered in [43], while such a network with multiple fixed RSs is studied in [8]. In both [8] and [43], time-division based half duplex transmission and AMC are assumed. In [43], it is required that WSs feed back their CSI to the BS or the RS in the form of per-subchannel SNR every time frame. Routing is decoupled from resource allocation to reduce CSI overhead, and an SNR-based path selection algorithm is executed to determine the node to route the data of a WS, an approach that has been previously used for non-OFDMA relay networks [37]. Since the BS-RS subchannels are assumed to have the same average SNR, only the number of required subchannels is estimated for relaying.

In [43] two algorithms (fixed time-division and adaptive time-division) are proposed to improve the cell throughput and coverage while minimizing complexity and overhead requirements, the frame is divided into a BS subframe followed by a RS subframe. In the first fixed time-division algorithm, the RS performs its own allocation in the RS frame and requests relayed users' packets to be forwarded by the BS. The RS, provided it has packets queued, allocates the channel to the WS with the highest SNR. Here, throughput is improved from exploiting traffic diversity gains. The set

of users with high SNR values and no queued packets are expected to be allocated in the next frame. Therefore, the RS requests the users' packets to be forwarded. The BS divides its resources between the RS's and the direct users' demand. The throughput within the time frame is defined as the average of the BS-RS throughput and the RS-WS throughput.

In the second adaptive time-division algorithm, the allocation is done just as in the first one. However, the time frame is quantized to an integer number of time slots and an iterative time-division adaptation is performed by adding or subtracting a time slot from the BS frame, to maximize the total throughput, which is averaged over the two hops. SNR-based path selection routing is considered, instead of joint routing and scheduling, to reduce the complexity of the algorithm and overhead. Furthermore, the time-division resource partitioning between BS and RS transmissions in a frame remains suboptimal despite the gain obtained by optimizing the duration of the subframes.

The objective function in [8] is the total average throughput of both the direct and relayed links presented as a function of SNRs and some indicator (optimization) variables. Two algorithms are proposed to improve the overall cell-throughput while minimizing the system complexity. In both algorithms the BS transmission frame is followed by an equal RS frame. The first algorithm performs subcarrier allocation with a predetermined equal power allocation (the same level for both BS and RS). The second algorithm achieves an optimal joint power and subcarrier allocation. Simulation results show that as the number of RSs increases, the sum rate is increased, while the joint allocation algorithm continues to outperform the fixed power allocation algorithm. Without relays, the joint algorithm introduces a marginal gain especially at high SINR regime, and the power allocation part becomes equivalent to water-filling.

In [9], an OFDMA relay network with multiple sources, multiple relays, and a

single destination is investigated where resource allocation is considered with fairness (load balancing) constraints on relay nodes. The authors' approach is to transform the integer optimization problem into a linear distribution problem in a directed graph to allow the use of the linear optimal distribution algorithms available in the literature. A (central) decision maker is assumed to know the CSI as well as the role of the node (source/relay/destination). Due to the low speed of nodes, this role designation does not change in an allocation period. The required information is collected at the central unit for allocation decisions. It is also assumed that the subcarriers allocated in the first hop are the same as those in the second hop. This reduction in available frequency diversity gain results in a potential system performance loss. Besides, centralized allocation schemes have an unenviable high signaling overhead problems. Thus, in dense relay networks, high complexity and signaling overhead requirements make them unattractive.

In [44] and [45], a multihop OFDMA-based downlink system is considered in a single isolated cell, based on a fairness-aware adaptive resource allocation scheme. The CSI is assumed to be available at the BS. The scheme performs adaptive subchannel, path, and power allocation to maximize the system capacity with minimum resources allocated to each user. The network architecture consists of a BS surrounded by three relays. Dedicated frequency bands are reserved for the BS-RS links, i.e., the resource allocation scheme considers only the links to users, either from the BS or RSs. The problem is formulated in a way that a WS can be assigned to either a BS or any of its RSs. However, as a result of excluding the BS-RS links from the resource allocation, a RS is assumed to always have data buffered for all of the potential users that can be served by this RS. WSs are allowed to receive allocation signals simultaneously from different nodes and to use the broadcasted resource allocation information to filter out received data.

This formulation leads to a mixed non-linear integer programming optimization

problem, which is known to be very difficult to solve in real time. Therefore, a best-effort fairness constraint is imposed to ensure that any user will be allocated a fixed minimum number of subchannels α where $0 < \alpha < N/K$, where N and K are the number of subchannels and users, respectively. The parameter α can be tuned such that the optimization problem results in strict fairness (as $\alpha \rightarrow N/K$) or strict capacity maximization (as $\alpha \rightarrow 0$).

The heuristic fairness-aware algorithm is composed of three steps; subchannel allocation, load-balancing, and power-distribution. The subchannel allocation step is realized in two phases. In the first phase, subchannels are allocated fairly by sorting all the channel gains (over all links and subchannels) in a descending order and allocating subchannels in that order. Once a user is allocated with its α subchannels the user is removed from the list. The remaining subchannels are allocated in the second phase to their respective best users to improve the capacity in a manner to achieve load-balancing which is to ensure that all nodes are allocated about the same number of subchannels based on the believe that the offered load at any node is proportional to the number of its assigned subchannels. The BS runs the load-balancing algorithm until the difference in the number of allocated subchannels for any two nodes is less than a chosen threshold, or when the difference in the channel gains assigned to different nodes on a selected subchannel is larger than another threshold. These two thresholds can be used to control the number of iterations. In the last step, the BS and RSs distribute the total transmission power equally to the allocated subchannels.

The performance of the capacity maximization heuristic algorithm is close to that of the optimal solution, and significantly outperforms the non-adaptive FDMA resource allocation scheme. In addition, at the price of loss in capacity, the fairness-constrained heuristic algorithm significantly increases the fairness index $F(t)$, based

on Jain's index [47], defined as

$$F(t) = \frac{\left(\sum_{i=1}^K r_i(t)\right)^2}{K \sum_{i=1}^K r_i^2(t)}, \quad (2.2)$$

where $r_i(t)$'s are user instantaneous rates. More elaborate discussion on Jain's fairness index is contained in Section 2.8.2.

To avoid a difficult non-convex optimization problem, [46] formulates a convex problem in a downlink single-cell scenario by using preassigned subchannels and optimizing only the power allocation. Thereafter, a heuristic scheme for subchannel-allocation using Lagrange dual-decomposition method is adopted. The authors propose a modified water-filling algorithm that can be solved by an inner-outer bisection method. In the half-duplex relaying scheme, the total transmission time frame is divided equally into two consecutive subframes, BS subframe and RS subframe. In the BS subframe, the BS transmits directly to the single-hop users on the set of subchannels B_0 and transmits to each of the M fixed relays on some other disjoint subchannel sets: B_1, B_2, \dots, B_M . In the RS subframe, RSs transmit to their respective users on the disjoint subchannel sets: S_1, S_2, \dots, S_M . The power constraint is applied to the problem in the form of total cell power constraint.

The optimal power allocation algorithm utilizes the Lagrange dual-decomposition instead of a subgradient method, which requires a large number of iterations to converge to the optimal solution. The optimal solution that satisfies Karush-Kuhn-Tucker (KKT) conditions implies that for any RS_j the sum-rate of the first hop over the set $B_j, j \neq 0$, and the sum-rate of the second hop over the set S_j must be equal. The solution finds a water-filling level k_1 for all the subchannels in the BS set B_0 and a pair of filling levels (k_{j2} and k_{j3}) for each RS_j , corresponding to the subchannels in the first and second hop sets, B_j and S_j , respectively. The following relation holds

for all j :

$$k_1 = k_{j2} + k_{j3}. \quad (2.3)$$

The proposed modified water-filling algorithm employs an inner-outer bisection method, where the outer iterations determine an upper bound of k_1 by applying the conventional water-filling on the subchannels in B_0 , while the inner iterations adjusts $\{k_{j2}\}$ and $\{k_{j3}\}$ to achieve the equality of the sum-rates of the two hops of all RSs. This equalization is realized in the three-step heuristic algorithm as follows:

Step 1: Generate the modified inverse subchannel SNRs for all $\{S_j\}$, i.e., $1/\gamma_1 + 1/\gamma_2$ where γ_1 and γ_2 are the normalized SNRs on the first and second hops, respectively.

Step 2: Determine k_1 using conventional water-filling on the set B_0 and the modified subchannels in $\{S_j\}$ under the total cell power constraint.

Step 3: Calculate the water-filling power required on the subchannels in $\{B_j\}$ using a bisection method.

The heuristic subchannel-allocation reads thus: the user with the highest SNR occupies subchannels in B_0 , and the user with the lowest modified inverse subchannel SNR occupies subchannels in $\{S_j\}$, whereas for $\{B_j\}$ the worst N_j subchannels of the BS are chosen.

The inference from this study can be summarized as follows: First, the subchannel allocation is a critical factor in the system performance which requires further investigations for finding an optimal scheme. Second, multihop relaying, along with the proposed heuristic subchannel allocation outperforms the optimal resource allocation in OFDMA networks without relays. The proposed optimal solution and the heuristic scheme have not efficiently exploited the inherent multiuser diversity. In addition, the schemes overlook the routing and scheduling in the optimization process. Because the treatment is based on simple network model, there are no provisions for how such schemes can be applied to OFDMA-based multicellular networks. This

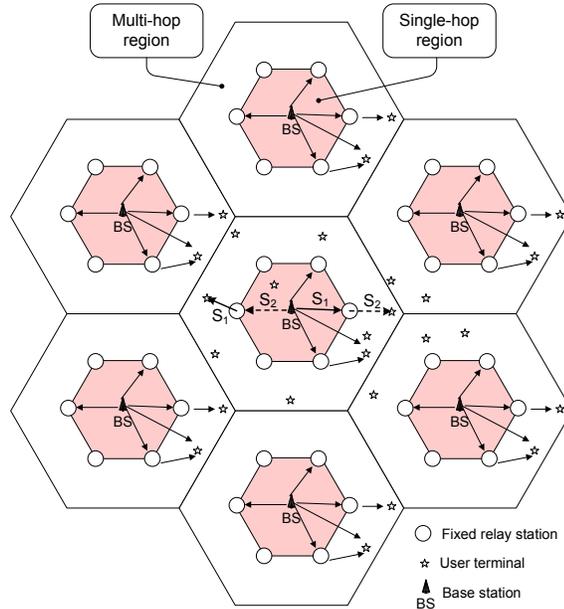


Figure 2.3: Network layout and spatial reuse pattern [48].

last point is applicable to many of the current literature, only a handful of papers are available in the open literature investigating relay-enhanced, OFDMA-based air interface technology in multicellular networks.

2.4.2 Centralized RRM Schemes in Multi-cell OFDMA Relay Networks

In [48], a centralized downlink OFDMA scenario in a multicellular network enhanced with six fixed relays per cell is considered. The proposed scheme considers efficient use of subcarriers via opportunistic spatial reuse within the same cell, as shown in Fig. 2.3 (i.e., a set of subcarriers used in a BS - RS link (S_2) can be reused after 180 degrees angular spacing in a RS - WS link), even when no directional antennas are employed. The data of WSs can be routed through any of the six relays, based on the maximum power received rather than the SINR. This means that the relay selection strategy is insensitive to the amount of interference received. The work in [48] is probably one of the first papers to consider such spatial reuse in a multicell environment, although, with a largely oversimplified model. Fading has not been considered except

for independent lognormal shadowing on links, which means that subcarriers are similar on any particular link. Consequently, the problem was formulated as the minimum number of required subcarriers to satisfy a user’s quality of service (QoS). This is used in the transmit scheme selection algorithm (TSSA) that switches among single-hop (SH), multihop (MH) and MH with spatial reuse (MHSR).

In TSSA, the only scheme allowed in the interior hexagon is the SH directly from the BS, whereas outside this region (relay region), TSSA could choose between MH scheme or MHSR that requires less number of subcarriers to satisfy the QoS requirements. With this strategy, an integer programming optimization problem is formulated to maximize the number of users with satisfied QoS requirements.

Using the percentage of unsupported users as a performance metric, the performances of TSSA and three other preset combinations represented as inner/outer region; SH/MH, SH/MHSR and SH/SH are compared. It is observed that a significant increase in the number of supported users is achieved when applying TSSA while the other three combinations performed in the following order: SH/MH, SH/MHSR and SH/SH. Restricting a cell region to a particular transmission scheme, regardless of the channel conditions, is suboptimal and potentially reduces the performance gains. In addition, no technique was employed to mitigate the co-channel interference. By ignoring channel variations and traffic diversity opportunities, the proposed scheme has not exploited the inherent multiuser and frequency diversity gains.

The idea of grouping users based on their locations for allocation purposes appears in [49] in the downlink OFDMA-based multicellular relay-assisted network with global frequency reuse. However, routing is decoupled from resource allocation. In the distance-based relay selection strategy, users within a predefined neighbourhood of the BS are restricted to a single-hop transmission from the serving BS. These users are called “near-users”, while the “far-users” are those that can only receive in two-hop transmissions through the closest RS. Two different time allocation policies are

considered in the cell capacity formulations. The first policy is the time-orthogonal policy, in which there is no frequency reuse at all on the three links; the BS-RS backhaul link, the RS-far user link, and the BS-near user link. These links are activated for only a fraction of the total time, μ_b , μ_f , and μ_n , respectively. In the second policy, on a common subchannel, the BS-RS backhaul link is active within μ_b of the total transmission time, while the BS-near user and RS-far user links are active simultaneously during the remaining $\mu_n + \mu_f$ portion. Intra-cell interference could occur during the latter concurrent transmissions, as no coordination is employed between the BS and the RS.

Of particular importance is the spectrum reuse capability in this scheme. The RS-far subchannel can be reused in the same cell by introducing a relay reuse factor such that all or a subset of the M RSs in a cell transmit in the time and frequency resource allocated to the BS-near user or each RS has to transmit on a subchannel orthogonal to the subchannels of the other RSs. In the latter case, there is no intra-cell reuse rather bandwidth is partitioned into M orthogonal blocks. In addition to multiuser diversity gains, further capacity improvement is realized by the spectrum reuse among the BS and RSs. It is observed that the performance of the proposed schemes is insensitive to the relay reuse factor, which means that such reuse does not improve the total capacity. The paper concludes that multihop relaying represents a viable method for improving the QoS for the far users.

It might have been obvious that in the future wireless networks, centralized RRM schemes are not the best option, considering latency, overhead, system and computational complexity, among other issues. Hence, the importance of distributed schemes has been recognized. However, there has not been much progress in this front.

2.4.3 Distributed RRM Schemes in OFDMA Relay Networks

In [50], a “semi-distributed” downlink OFDMA scheme in a single cell enhanced by M half-duplex fixed relays is considered. The scheme divides the users into disjoint sets located in the neighbourhoods of the BS and RSs, the approach that is common in the literature and discussed above. The users attached to the BS and relays are referred to as the BS-WS cluster and the RS-WS clusters, respectively. The BS allocates some resources to the BS-WS cluster directly and to the RS-WS cluster through the RS. Implicitly, it is assumed that all routes have been established prior to resource allocation, regardless of the channel conditions, and that the same subcarrier is used on the two hops, BS-RS and RS-WS. In addition, it is assumed that a protocol is available for gathering CSI and the allocation decisions are broadcasted on a separate control channels.

The authors in [50] classified their algorithms into *Separate and sequential allocation (SSA)* and *separate and reuse allocation (SRA)*. The starting point of these two-step resource allocation schemes is basically the same. In this first step, each RS, along with its user cluster, is treated as a large-sized WS with a required minimum rate equal to the sum of all the minimum required rates of the WSs in its cluster. The CSI, in terms of received SNR, is fed back from the RS in the form of two $(1\text{-by-}N)$ vectors carrying the received SNRs at the RS itself and a processed version of the received SNRs at its connected WSs, on N subcarriers. The processing function can result in the maximum SNR, the minimum SNR or the average SNR on each subcarrier. Hence, the BS allocates the resources among its own WSs and these virtual large-sized WSs. In the second step, the RS allocates resources to the users in its cluster based on one of two allocation schemes:

- Resources assigned in the first step to that BS-RS link are allocated among the connected users, this is the SSA.

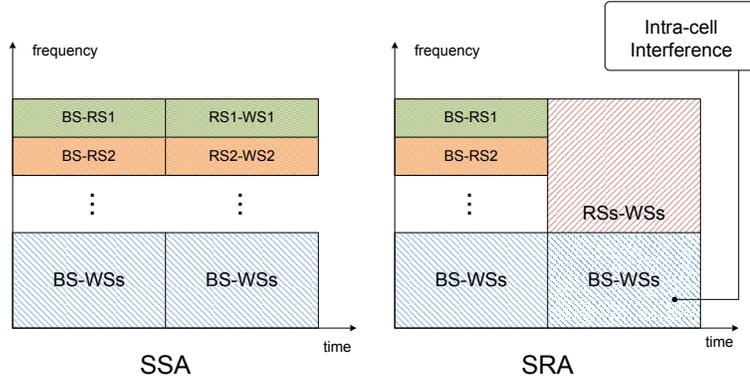


Figure 2.4: Half duplex relaying patterns for SSA and SRA as in [50].

- The RS re-allocates all the N subcarriers to its connected users regardless of the BS assignments, this is the SRA.

Figure 2.4 shows the difference in relaying between SSA and SRA within two equal and consecutive time frames. Since both RS and BS could assign the same subcarriers to their respective users, SRA is prone to intra-cell interference

The SSA and SRA are compared with the centralized resource allocation scheme. In the three schemes the best user on any subcarrier is assigned at any decision-making node, a continuous rate adaptive modulation is employed as in (5.1). Simulation results for a single cell with one relay show that the semi-distributed scheme, SSA in particular, has a comparable capacity and outage probability performance to the centralized scheme. The SSA shows significant performance stability over the SRA. Intra-cell interference that could occur during the RS frame in the SRA scheme brings considerable increase in outage probability. The allocation at the RS spans all the N subcarriers and has no coordination at all with that of the BS. Thus, the system could suffer from instability.

In general, the proposed semi-distributed schemes reduce the amount of overhead required to feed back the CSI and minimum rates to the BS, but in the case of SRA, we observed that there is no need for communicating such information to the BS. These schemes fail to exploit the interference avoidance and traffic diversity gains. In

addition, there is an inherent loss in performance due to the decoupling of routing and scheduling processes. An excellent tutorial outlining the promises in well-designed distributed RRM schemes is available in [51].

2.5 RRM Opportunities and Challenges in OFDMA Cellular Relay Networks

The focus of this section is mostly on the potential RRM opportunities and challenges associated with the migration of OFDMA networks from the conventional cellular architecture to the relay-enhanced. To be on the side of caution, it may be necessary to re-examine the technical terminologies used in conventional networks as they relate to this new paradigm shift, i.e., relay-enhanced networks.

2.5.1 Migration from Conventional Cellular to Relay-Enhanced

2.5.1.1 Centralized or Distributed?

There seems to be no agreement in the literature on the use of the terms “*centralized*”, “*decentralized*”, and “*distributed*” as regards to the operation of RRM schemes. For instance, in the context of conventional cellular networks (e.g., [51]), an RRM scheme is considered to be centralized if there exists a central controller that gathers all the information and feedback required from all BSs and performs global resource allocation as shown in Fig. 2.5. Whereas, an RRM scheme may be considered distributed if each cell performs its own resource allocation individually based on local information, and perhaps, aided with some inter-cell information. In the context of relay-enhanced cellular networks, however, the latter scheme is referred to as centralized given that the BS gathers information from the RSs and performs the resource allocation for itself and the RSs as well.

One way to avoid such ambiguity is to tie the description to the relevant network

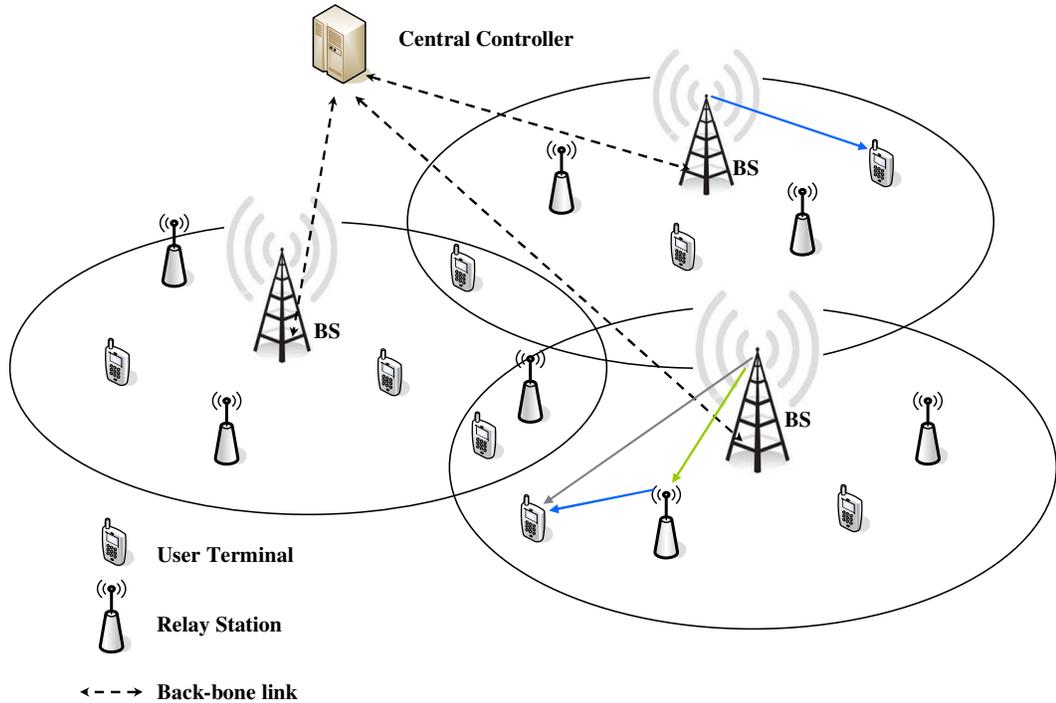


Figure 2.5: Example relay-enhanced cellular architecture. A central controller performs global resource allocation in centralized conventional networks.

“hierarchy”. For instance, a network-level distributed/cell-level centralized RRM scheme could refer to a scheme that does not rely on a central controller to perform global resource allocation while each BS will individually handle the resource allocation for all the entities in the cell including RSs. Similarly, a network-level distributed/cell-level semi-distributed RRM scheme could refer to a similar scheme but with the provision that RSs can partially participate in the resource allocation. The importance of such precise terminology cannot be overemphasized in the future networks where several entities will be involved, in one way or another, in the realization of RRM schemes.

In addition to that, the ambiguity of the technical terminologies extends to an important aspect which is “load balancing”.

2.5.1.2 Cell-load Balancing

Load balancing is a function usually incorporated with the connection admission control (CAC) mechanisms in conventional cellular networks. In that context, the load balancing function refers to the handover (hand off) of some users between adjacent cells to distribute the traffic load network-wide among BSs while maintaining users' quality of service (QoS). Although the load balancing as defined above will be an integral part of any prospective RRM scheme in the relay-enhanced networks, researchers often associate the term "load balancing" in relay networks with a different function which aims at distributing the load evenly among all nodes, cell-wide. The number of OFDM subcarriers handled by a node is often employed in literature as a good estimate of the traffic load at that node [44], [52], and [53]. A balanced traffic load reduces the packet processing delays at the regenerative relays.

One might refer to the first definition as *network-load balancing* while the second definition might be referred to as *cell-load balancing*. In fact, cell-load balancing also has some implications on inter-cell interference and fairness which we highlight in Subsection 2.5.2 and Section 2.8, respectively. We will discuss the concept of network-load balancing in relay networks in Subsection 2.5.1.5.

2.5.1.3 Various Forms of Wireless Relays of Different functionalities

It is envisaged that a plethora of relay stations of different specifications, functionalities, and geographical densities, will be part of the next-generation cellular network architecture. As such, the prospective RRM schemes should be able to distinguish between these different types of wireless relays and exploit their unique characteristics. Moreover, strategies for handling the heterogenous relay network resulting from the coexistence of these relay types will have to be considered. For instance, fixed relays are assumed to be deployed with low density at strategic locations of the cell;

possibly line-of-sight communication with the BS is maintained and wall outlet power is available. This is totally in contrast to the mobile relays for which the channel, protocols and power budget are all characterized by their mobility, e.g., roof-top vehicular devices. Another different device is the autonomous (plug-and-play) type of relay known as a nomadic relay which is a portable battery-powered device and mostly deployed by the users. In [54], motivational scenarios for using mobile multihop relaying are provided with emphasis on the use cases of these types of wireless relays. On that note, the first part of this thesis considers fixed-relay cellular networks whereas the second part considers the integration of a high density of nomadic relays into the fixed-relay network.

As the architecture of next-generation cellular networks becomes more sophisticated comprising a plethora of active nodes, distributed RRM schemes with limited feedback, especially if involving mobile relays, become essential. Furthermore, in a network with a large number of relays, more often than not, orthogonal resources are required for multihop relaying purposes and thus a form of reuse is necessary. To facilitate this reuse and combat the resulting interference, intelligent RRM schemes are needed to balance between aggressive resource reuse and efficient management of the associated co-channel interference (CCI).

Given such modern architecture, prospective RRM schemes are quite diverse in terms of transmission protocols and optimization objectives. Among the fundamental questions these protocols have to answer are the following; which transmission mode (direct, multihop simple relaying, or multihop cooperative relaying) is optimal for a particular user and which relay node(s) should be incorporated in that mode. This is where efficient routing schemes come as an RRM design tool.

2.5.1.4 Relay Selection or In-Cell Routing

Routing is thus a key issue of networks that support multihop relaying through deployment of dedicated relays, users' cooperative relaying, or protocols incorporating both. Routing can be viewed as the process of establishing efficient connectivity between nodes over multihop links allowing coverage extension, throughput, and fairness improvement. Since different routing schemes are expected to affect the system performance differently in terms of throughput, delays, and signaling overhead, several relay-selection strategies and relaying criterion are employed in the RRM schemes as an initial step followed by scheduling user packets on the chosen path(s), e.g., [48]. However, performing routing and scheduling jointly is known to produce superior performance results as compared to decoupled scheduling and routing [38]. Therefore, expressed differently, resource allocation in such relay-enhanced networks is indeed a joint scheduling and routing problem. However, it is quite challenging to devise efficient RRM schemes that tackle the joint problem. The algorithm presented in Chapter 3 is a good example though of a class of dynamic joint routing and scheduling strategies discussed in literature [55], based on the theory in [56], as applied to OFDMA relay networks [57]. In such a strategy, CCI is depicted by the achievable rates of individual hops on different OFDM tones.

In addition to the multihop relaying schemes, the joint routing and scheduling problem challenges as well the multihop cooperative schemes in relay networks. In fact, the authors in [58] observe that in order to devise efficient routing algorithms to establish a path through multiple clusters of possible cooperating nodes, the problem is in principle joint routing, clustering, and resource allocation.

It is worth mentioning that in the context of relay-enhanced cellular networks, 'in-cell routing', or simply relay selection, in case of two-hop relaying within the cell vicinity, is the most commonly considered [37]. Nevertheless, more sophisticated

schemes that enable a sort of mesh topology exploit the substantial increase in degrees of freedom by establishing routes comprising relay nodes even located in adjacent cells, i.e., inter-cell routing. Such RRM schemes, despite incurring further complexity, are of great research interest [59].

2.5.1.5 CAC and Handover for Network-Load Balancing via Inter-Cell Routing

As we mentioned earlier, RRM schemes have to work in conjunction with a CAC mechanism [60], which decides, based on available resources and connected users' QoS, whether to admit an incoming connection to a particular cell (BS) or deny it and handover the user to a neighboring non-congested cell through a handover mechanism. Such mechanisms are essential to balance the load network-wide and reduce the blocking probability. With the deployment of relays, more handover opportunities arise through enabling inter-cell routing. In that case, a user with denied connection to one cell can be admitted to an adjacent cell by establishing a connection through one or more RS(s) in the latter. This dynamic load balancing mechanism is termed 'primary relaying' in [61], in which the authors integrate ad hoc relaying schemes into cellular networks. An alternative mechanism termed 'secondary relaying' is also proposed where an ongoing connection can be diverted through RSs to an adjacent cell and the vacant resources are then inherited by the incoming connection. Such a mechanism is beneficial whenever handover opportunities are limited for the user with an incoming connection due to traffic and channel conditions.

Although the concept could be generalized, to alleviate the burden on the cellular resources, these dynamic load balancing mechanisms rely on the availability of out-of-band channels, such as the 2.4-GHz Industrial, Scientific, and Medical (ISM) band, to facilitate multihop relaying and user access [62]. Several works have analyzed the performance of these mechanisms and proposed different routing or path selection

criteria; a mathematical theory has been developed in [63]. It has been noted however in [58] that out-of-band relaying requires UTs and RSs equipped with multiple radios operating on different frequency bands; the default cellular interface and an ad hoc wireless interface. Moreover, we observe that these inter-cell routing mechanisms were proposed and analyzed for TDMA/CDMA-based networks rather than OFDMA-based relay networks.

Mesh routers employing a combination of WLAN and OFDMA-based WMAN radios have been considered though in [64] where a game-theoretic framework for bandwidth management and CAC in an integrated WLAN/WiMax multihop relay architecture is proposed to provide service to mobile hotspots. The admission control criteria therein limits the number of ongoing connections at a mesh router so that their total utility is maximized.

2.5.2 Co-Channel Interference in OFDMA Cellular Relay Networks

Another burdensome challenge that faces the RRM schemes in next-generation cellular relay networks is the increase in *co-channel interference* (CCI). In fact, CCI is inherent in any multicellular network mainly due to inter-cell and intra-cell resource reuse. Clustering, cell sectorization, and static/dynamic fractional frequency reuse are common techniques used to control the inter-cell interference (ICI) in conventional cellular networks. With the increasing demand of high data rates in next-generation networks, highly aggressive reuse schemes are envisioned to achieve a much higher spectral efficiency. The aggressive reuse suggests that entire system resources will be made available in each cell while intra-cell spatial reuse can be applied to further improve resource utilization given that a part is consumed in facilitating multihop relaying on orthogonal channels.

On the cell level, some techniques such as dynamic power allocation [55] can be employed, as a link adaptive technique, to either mitigate or cope with the excessive

CCI in such interference-limited systems. It is however observed that the vast majority of works apply static intra-cell spatial reuse patterns based solely on user locations, e.g., reusing the channels assigned to RS-UT links in the BS-UT links within the close vicinity of the BS [43], [49]. Such static reuse patterns are not informed by the resultant CCI. However, more intelligent RRM schemes are supposed to control the CCI through opportunistic intra-cell reuse utilizing instantaneous channel conditions and antenna directivity [65]. In such cases, least CCI levels are attained, under-utilized resources, if any, are used first. Thus, different channels will have different reuse factor realizations.

More importantly, while relays are deployed with the potential of improving coverage and assisting users having unfavorable channel conditions (e.g., cell-edge users), an adverse effect arises. That is, in the downlink scenario, relays deployed in one cell bring the interference closer to the cell-edge users in the adjacent cells; this potentially increases the level of inter-cell interference and renders a more interference-limited system for the prospective RRM schemes in their attempt to attain the desired high spectral efficiency.

Therefore, some recent works, e.g., [66] and [67], have extended the static fractional frequency reuse technique to OFDM cellular relay networks as a compromise solution between aggressive reuse and cell-edge performance due to the inter-cell interference caused by the neighboring relays. In both works, a three-sector cellular network is considered where the fractional frequency bands are used for the RS-UT communication at the relay coverage area while the whole band is used for the BS-UT and BS-RS communication in each sector. However, such techniques require planning and limit the opportunities a prospective relay-based scheme could exploit in frequency, multi-user, and spatial diversity within the relay coverage area.

Having observed that cell-edge relays potentially increase the level of ICI, an interesting question could be; can relays help mitigate the inter-cell interference? In

fact, some works such as [68] and [57] do employ the cell-edge relays in mitigating the inter-cell interference. In [68], an RS located at the cell-edge is shared by the surrounding BSs and is equipped with multiple antennas. The shared RS separates the received mixed signals and performs *interference suppression* using MIMO techniques, on the same resource block, before forwarding to respective UTs; we will revisit this approach in Section 2.6. Using a different approach, the algorithm in [57] (discussed in the following chapter) utilizes the existence of relays and their uniform geographical deployment to spatially *randomize the ICI* through its inherent cell-load balancing feature which results in an even distribution of OFDM subchannels among the active relays of each cell. As such, a cell-edge user being served by a relay in one cell is less likely to receive ICI from the closest relay of the adjacent cell.

On the network level, *dynamic inter-cell coordination* could be the candidate for addressing interference problems. Dynamic inter-cell coordination can be employed by exchanging vital interference information among BSs over the backbone network connection to achieve the interference avoidance gain and improve user throughput and/or fairness. Although several inter-cell coordination schemes have been proposed in the literature for conventional and LTE cellular networks [69], [70], there have not been proposals for dynamic coordination schemes designed for cellular relay networks so far. Interestingly, our efforts towards devising dynamic inter-cell coordination schemes for OFDMA-based cellular relay networks led to some important observations on the issue of pronounced ‘interference uncertainty’.

The challenges of Co-Channel Interference Uncertainty: Due to the dynamic and synchronous nature of the network-level distributed resource allocation process in multi-carrier cellular networks, the lack of interference predictability represents a challenging problem that has been observed in the literature of conventional OFDMA-based cellular networks. Generally, such uncertainty of subcarrier quality arises when the network is partially loaded such that BSs, based on individual allocation decisions,

are not fully utilizing the set of available subcarriers. Various techniques that aim at increasing the predictability of interference in conventional OFDMA-based cellular networks are discussed in [51] which we summarize as follows:

- Partitioning (imposing structures): Power shaping over time slots within the frame, on-off power shaping over the cell sectors in time, fixing the sequence of resource allocation to users, or limited exchange of interference information between neighboring BSs.
- Discretization of the quantities involved in the resource allocation such as transmit power levels, steering coefficients of the directional antennas, and transmission rates.

Although some of these techniques may still be imported to relay-enhanced type of networks, they will basically compromise the aggressiveness in resource reuse and limit the available degrees of freedom. Moreover, these techniques are anticipated to be less efficient in combatting the uncertainty problem as it becomes more pronounced in relay networks. The main reason is not far-fetched; the problem originates from the dramatic increase in the ICI dynamics resulting from switching over different link budgets of potential interfering links as the subcarrier assignment hops among different cell nodes (see Fig. 2.6).

In other words, a receiving node on a particular subcarrier, somewhere in the network, will experience a dramatic change in interfering signal strength when the subcarrier-to-node assignment changes in the interfering cell from an allocation instant to another. To the best of our knowledge, current literature on relay-enhanced networks overlooked such deterioration in the CCI uncertainty. However, our results for RRA schemes and some dynamic inter-cell coordination schemes as applied to OFDMA cellular relay networks revealed the following:

1. Some dynamic inter-cell coordination schemes may not tolerate the increased

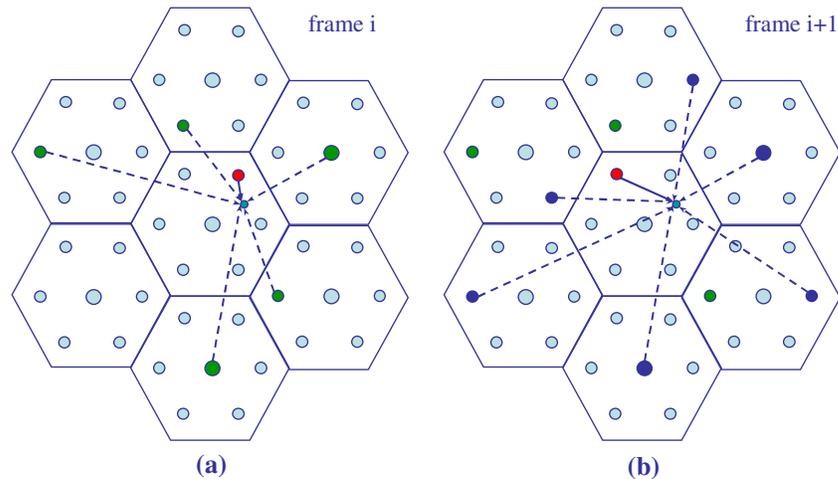


Figure 2.6: Snap shots of the cellular network in downlink explaining the CCI uncertainty problem: (a) CCI during frame i , (b) CCI during frame $i + 1$. Solid arrows represent desired signals in the center cell.

CCI uncertainty as their embedded optimization criteria rely on the assessment of how much ‘harm’ each individual interferer causes to other nodes before taking decisions against the most harmful interferers. As such, relying on the outdated interference power observed during the previous transmission can obviously overestimate a tolerable interferer or underestimate a very strong one in the following transmission. Therefore, the overall performance gain in throughput or fairness might be insignificant, if not in the negative direction. For instance, this has been observed in the performance of a binary power control coordination scheme in which a central controller is employed to only resolve the conflicts in suppression requests issued by the receiving users or relays against their most harmful interferers in the adjacent cells.

2. On the other hand, RRA algorithms designed for OFDMA-based relay networks might, however, tolerate the increased uncertainty as the subcarrier quality metric used is usually tied to the sum of received interference power, e.g., signal-to-interference-plus-noise ratio (SINR), rather than individual interferer strength. Thereafter, if practical adaptive modulation and coding (AMC) is employed,

as a link adaptive technique, a wide range of these SINRs is then quantized to one out of a few AMC modes or spectral efficiencies [71]. More importantly, any greedy or even fairness-aware RRA algorithm, will end up allocating the ‘best’ subcarrier through a comparative or a sorting routine. As such, the allocation result is not necessarily sensitive to the actual values of the outdated quality metrics or SINRs (observed during the previous transmission) and therefore the algorithm might still achieve its desired objective without significant performance losses.

Since the last observed interference power values may not represent the actual interference situation when transmission takes place, optimization could rely on a statistical value of the past sequence of previous interference measurements. For instance, a moving-average interference filtering approach has been examined in our studies for its simplicity in terms of computation (similar to the filtering used in the proportional fair scheduler). This recursive single-tap filter basically provides one output sample $\tilde{I}(t)$ of interference that is more influenced by immediate past measurements than earlier ones, within a time window controlled by α , as

$$\tilde{I}(t) = (1 - \alpha)\tilde{I}(t - 1) + \alpha I(t - 1), \quad 0 < \alpha < 1. \quad (2.4)$$

The parameter α can be carefully chosen. A review on the research development towards adapting the transmission using the statistics of measured SINRs, rather than relying on the outdated information itself, is provided in [72].

2.5.3 Restating the Strategy for Improving the Goodput Efficiency

As mentioned earlier, both dynamic inter-cell coordination and link adaptation are possible techniques for CCI avoidance or mitigation. Both techniques rely on *channel state information* (CSI) feedback from the receiver to the transmitter, as frequent as

the relevant resource allocation process. Since the net throughput or ‘goodput’ depends on how much feedback overhead penalty the RRM scheme incurs, the CSI feedback represents an additional challenge to any prospective centralized RRM schemes in OFDMA-based cellular relay networks. This is because, in principle, the required CSI feedback in such networks is substantially higher than that required for the conventional cellular networks due to the following reasons:

- Signals at source nodes have multiple potential recipients among which the routing algorithm selects the appropriate receiving nodes.
- Relays, as receivers, are also required to provide CSI feedback to the source node(s).
- Half-duplex relaying is commonly employed due to current practical limitations on the operation of wireless relays. Therein, the transmission time frame is split into two consecutive sub-frames (e.g., in downlink, BSs transmit and then BSs and RSs transmit) and thus links that are active during both sub-frames experience two different interference situations which need to be reported within the duration of a frame as compared to a single interference situation in case of conventional cellular networks. In fact, this applies even if the channel gains do not change along the whole frame duration.

2.6 Role of Multi-antenna Systems in the Future Networks

The combination of multi-antenna systems in the form of multiple-input multiple-output (MIMO) and OFDMA has been shown to yield a rich synergy. From RRM perspective, MIMO in its simplest implementation, can significantly improve the multiplexing gain of system links. Such boost in the quality of communication links yields significant savings in the system resources as less resource blocks are needed therein to satisfy the users’ QoS requirements. Thus system capacity in terms of the number of users and/or traffic load is increased and a reliable service is attained. In fact, The

LTE technology has positioned itself to exploit this opportunity. For conventional MIMO, different UTs are expected to carry multiple antennas according to their size constraints.

Different strategies are required in order to enjoy the benefit of MIMO transmission in a scenario where multiple antennas are not deployable at the terminal. For instance, cooperative relaying schemes and protocols are envisioned to build virtual antenna arrays from distributed single-antenna terminals [73]. This however is possible in the evolved LTE, i.e., LTE-Advanced where relays are an integral part of the technology. The current LTE specification indicates that mixed (in terms of the number of antennas) terminals are supportable and thus point-to-point MIMO can be realized for terminals equipped with multiple antennas. For the single-antenna terminals, the conventional maximal ratio combining technique can be used to improve the system reliability since the eNodeBs (LTE base stations) are deployed with multiple antennas.

Another important aspect of multi-antenna relay-aided systems is the opportunity of establishing efficient multihop routes through beam-forming and the smart antenna capabilities of nulling the CCI from neighboring users or relays. A more elaborate exploitation of the combined relaying and multi-antenna technologies is observed in [68] where multi-antenna cell-edge relays are shared by the surrounding BSs and separate the received mixed signals using MIMO techniques. Without resorting to resource planning and partitioning, such scheme alleviates the interference burden on system resources and further facilitates aggressive reuse.

In summary, multi-antenna systems will provide a great opportunity for advanced signal processing (such as beam-forming, pre-coding, and multiplexing, among others) essential for reliable and bandwidth-efficient system deployment.

2.7 Extending Conventional scheduling algorithms to Cellular Relay Networks

A common strategy to extend a conventional non-relaying scheduler to relay networks, though not optimal, is the generic framework of partitioning the users into clusters around the chosen serving nodes (BS and RSs), and based on this partitioning resources are shared among the nodes. Each node then schedules the users as in conventional OFDMA systems. Examples are the partial proportional fair (PPF) scheduler and the extended/greedy round-robin schedulers in [74], and the reference scheme in [75]. The proportional fair scheduler (PFS) is a widely used scheduler that provides a compromise solution between network capacity-greedy scheduling and user-fairness oriented scheduling. Thus, it realizes the multiuser diversity gains to some extent while maintaining a degree of fairness across UTs [76]. A simple implementation of this scheduler allocates a resource unit to the UT that maximizes the ratio of its achievable rate on that unit to its exponentially weighted average rate [77].

Although PFS is known in the literature to provide an efficient throughput-fairness tradeoff, incorporating this scheduler in multi-carrier systems has not been investigated vigorously [78]. This is going to change soon given the recent developments in the literature. Through the strategy described earlier, some heuristics have integrated PFS into combined relay and OFDMA technologies such as in [49], [79], and [80]. Since the BS node is required to allocate the resources among the direct UTs and the feeder links of the RSs, a priority metric for such feeders to contend with direct UTs has been proposed in [79]. For the relay-enhanced scheme proposed in [80], a potential improvement in proportional fairness sense can be realized through the clustering (in-cell routing) criterion of UTs which aims as well at maximizing the proportional fairness metric. PFS has been also implemented in [81] considering both half- and full-duplex UTs in a multihop FDD network. However, due mainly to the

inherent lack of queue-awareness in the PFS besides the suboptimal extension strategy, the relay-enhanced scheme proposed in [57] is shown to provide more fairness in throughput rate sense.

Beside the prominent challenges, it is important to point out the fairness opportunities residing in such rich environment with numerous degrees of freedom. In the following section, we discuss various fairness types, fairness assessment methods, as well as some possible fairness implementations in prospective RRA algorithms.

2.8 Fairness in OFDMA Relay Networks: An Obligation or A Privilege?

Schedulers in OFDMA-based networks can be designed to fully exploit multiuser and channel diversities in both time and frequency to maximize the total cell capacity in a greedy manner at the expense of fairness among users. Such a design approach is network-centric and does not take individual users' QoS into account. Thus, the applicability of these algorithms in practical cellular networks is doubtful since users pertaining to the same service class will be charged similarly while the service is not distributed evenly. Clearly, from a roaming user's perspective, the inability to maintain fairness defeats service reliability and ubiquity as it becomes channel and location dependent. That would be deemed a failure on the part of the service provider, despite the technology and ingenuity embedded in network equipment. In that sense, fairness as an obligation rather than a privilege ensures user satisfaction, regardless of location and channel conditions.

Therefore, user fairness expectations would be even higher in modern relay networks where service providers advertise outstanding quality of service based on the new architecture widely adopted by the state-of-the-art standards. With less diversity sacrifices in mind, relaying promises ubiquitous coverage and QoS improvement for

the users with unfavorable channel conditions. Fairness obligation is thus stronger, whereas keeping such promises is conditioned on the awareness and smartness of the RRM scheme in use.

Relative fairness [82] is a very important aspect when it comes to practical implementation. That is because users subscribe to different service classes offered by the service provider for different cost per bit. Thus, heterogenous traffic flows are expected to coexist with different statistical nature and QoS requirements. Therefore, fairness assessment has to be conducted within subgroups of users that pertain to the same service classes taking into account their respective on and off traffic bursts. Otherwise, in order to incorporate multiple classes into the fairness assessment, some works have proposed modified fairness metrics or assessment criterion by introducing relative weights [83] or using normalized throughput [84], [85]. In that sense, results reflect on the relative fairness yet they may vary from a time window to another within a session due to the different traffic properties.

Various fairness classes are noted in [31] such as short-term fairness, long-term fairness, and time-average fairness. The classification is dependent on the time window size used to evaluate the chosen fairness metric. Generally, achieving short-term fairness imposes stringent constraints on the RRM scheme while less stringent constraints for the long-term fairness and relaxed constraints for the time-average fairness. The choice of the appropriate fairness class to investigate the fairness of an RRM scheme should be based on the traffic model and the relevant QoS requirements.

Relay fairness is another important aspect of RRM in OFDMA-based relay networks which is different from user fairness because there are no QoS requirements specific to RSs. In fact, relay fairness as appearing in the literature, aims at distributing the traffic load almost evenly among RSs so that no RS will be overloaded [9]. In [86], relay fairness is assessed based on the power consumption at the relays, under different relay selection mechanisms, to operate a cooperative diversity scheme

without overloading the battery of one or more relay(s). We also note that if the relay's transmit power per subcarrier is fixed, maintaining an almost even distribution of subcarriers among relays limits each relay's total transmit power and thus its power amplifier rating and the consumption of its battery energy, for battery operated relays. In addition, a balanced traffic load reduces the packet processing delays at the regenerative relays. Even in the context of integrated WLAN/WiMAX multihop mesh networks considered in [64], it is observed that the network resource utilization can be improved by balancing and sharing the load among the mesh routers (relay nodes with multiple radios), through efficient routing mechanisms.

Therefore, we observe that the term 'cell-load balancing', as defined earlier in Section 2.5, and the term 'relay fairness' can be used interchangeably, regardless of the measure used to quantify the load.

2.8.1 Notion of Fairness in this Thesis

In the first part of the proposed work of this thesis, an OFDMA cellular relay network is considered where users pertaining to the same service class receive statistically symmetric inelastic traffic processes. Whereas, the statistics vary from a class to another. In order to achieve a ubiquitous and reliable service in such system, the RRA scheme has to dynamically route, and allocate appropriate resources to, each admitted user's traffic flow *regardless of the user's location, instantaneous traffic bursts, and short- and long-term channel conditions*. In other words, the scheme has to achieve throughput fairness within each class of symmetric traffic flows, and more importantly, achieve relative fairness across these asymmetric classes such that light traffic flows are not deprived resources due to the heavy (or temporarily heavy) traffic flows. These notions of fairness and ubiquity of throughput-optimal scheduling are quite uncommon in the literature due to the absence of a cellular system model and/or the lack of an efficient implementation that reveals such behavior.

It is also worth emphasizing that such fairness notion does not contradict our intuition of the throughput-fairness trade-off commonly observed in the literature that considers systems with continuous backlogs or full buffers. In fact, this full buffer scenario has been adopted in the second part of the thesis where we address the integration of self-organizing nomadic relays into the fixed-relay network. In full buffer models, the user with the highest achievable rates would always achieve the maximum resource utilization and thus the maximum throughput, if assigned the whole resources on the expense of fairness. In contrast, our RRM schemes in Chapter 5 maintain the ubiquity and reliability in terms of the minimum QoS for all users, whereas the schemes proceed thereafter in a greedy manner. It is clear though that the aggregate throughput in such case would be less than that of a pure greedy scheme due to that degree of fairness we achieve. We also note that the minimum QoS level can be tuned by the operator for each service class, if any.

2.8.2 Fairness Metric Examples

Fairness functions can be viewed from two different perspectives: Fairness criteria implemented in routing and scheduling algorithms to assess priorities of users and evaluate whether compensation is required; and fairness metrics used to investigate and classify fairness-aware algorithms. In the following, we briefly mention the most commonly used fairness functions or metrics

1. *The proportional fairness metric*; considering a network with K users with the same service class and priority, the scheme that maximizes the proportional fairness metric, expressed as

$$F = \sum_{i=1}^K \log r_i, \quad (2.5)$$

is more fair according to the game-theoretic definition of proportional fairness

which implies that any change in the ‘proportionally fair’ rate allocation \mathbf{x} , to say rate allocation \mathbf{y} , must have a negative total relative change [76], i.e., $\sum_{i=1}^K \frac{r_i^y - r_i^x}{r_i^x} \leq 0$.

2. *The Jain’s fairness index* [47] assumes a value between $1/K$ and unity for a network having K users with the same service class and priority. Mathematically, Jain’s index can be expressed as in (2.2). Figure 2.7 shows the behavior of Jain’s index for a 2-users case. The index possesses low sensitivity at high rates, where small variations between users can be neglected, and high sensitivity at low rates where similar variations are not negligible.
3. *The IEEE 802.16m fairness index* [87] maps each user’s rate to a value between 0 and K as expressed in the following:

$$x_j = \frac{r_j}{\frac{1}{K} \sum_{i=1}^K r_i}. \quad (2.6)$$

Figure 2.8 shows the behavior of the index expression for user 1, x_1 , in a 2-users case. It can be seen that the behavior of x_2 occupies the volume that is complementary to that of x_1 . Note that the function is not finite when all users achieve exactly zero rate. However, this is not a practical situation since the RRA algorithm would allocate, in the worst case, all or a portion of the resources to some user.

4. *The fairness factor* [31] has been developed to measure fairness in RRM noncooperative static games. As shown in (2.7), this factor represents the normalized statistical standard deviation of user’s throughput compared with that of a single-user case. In contrast to Jain’s index, the higher the factor the less the

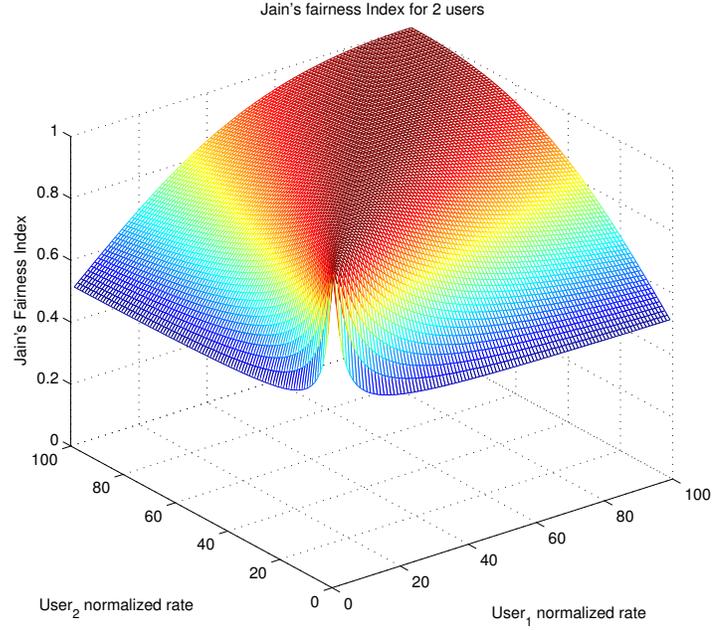


Figure 2.7: The behavior of Jain's index function in a 2-users case.

fairness realized among users.

$$\rho = \frac{1}{\bar{T}} \sqrt{\frac{1}{K-1} \sum_{i=1}^K \left(\frac{T_i}{T_i^{\max}} - \bar{T} \right)^2}, \quad (2.7)$$

where T_i^{\max} is the maximum throughput user i could achieve solely while $\bar{T} = \text{average}(\bar{T}_i)$, $\bar{T}_i = \text{average}(T_i/T_i^{\max})$.

2.8.3 Fairness Implementation in RRA Algorithms

Basic scheduling approaches and criterion (e.g., max-min or round-robin and proportional fairness) as well as the popular practical schedulers such as: the channel state dependent packet scheduling (CSDPS), the channel independent packet fair queueing (CIF-Q), and the server-based fair approach (SBFA) were proposed for the conventional networks. There are no provisions for how such approaches should be applied

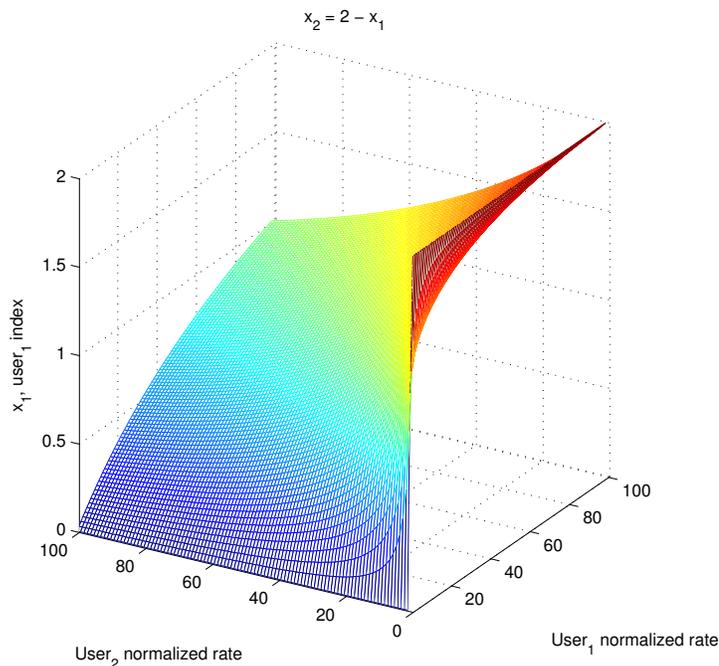


Figure 2.8: The behavior of the IEEE 802.16m index in a 2-users case.

to a relay-enhanced cellular network, which is different from the conventional down-link point-to-multi-point problem. Therefore, new algorithms need to be developed to facilitate fair scheduling and routing in such networks. While the literature in this area is still gathering momentum, one can realize that most of the algorithms proposed so far focus on total capacity maximization and, sometimes, with some fairness constraints imposed on the optimization problem, e.g., [44] and [8].

Since capacity does not map directly to throughput due to traffic burstness, such scheduling and routing algorithms will not satisfy the QoS requirements because they are unaware of traffic and queues status. In other words, allocating resources for some links based solely on SINRs or achievable rates, even under fairness constraints, can result in substantial performance losses if the buffers of source nodes contain insufficient data at that allocation instant. For that reason full queues at all potential source nodes, despite unrealistic [43], are often implied in some studies, e.g., [48].

An interesting approach is to involve buffer states or queue lengths in the formulation of the optimization problem and thus not only is wastage of resources avoided but also traffic diversity is exploited, which means that when some users' traffic is in the off period, more resources can be utilized to provide a better and more fair service to the other users.

One way of involving the buffer states in formulation is to work out the optimization problem as a sum-utility or sum-demand maximization where such metric is proportional to both the queue length at the source and the link quality at destination. In [88] and [89], modified versions of this metric are used in conventional cellular SDMA/TDMA and OFDMA networks, respectively. Generally, queue-awareness allows RRM schemes to take corrective actions in following allocation frames to compensate the overlooked user buffers, if any, and potentially improve fairness in the long term. However, these corrective actions are limited to subchannel allocation in the conventional RRM schemes which clearly can not combat large pathloss unless inter-cell routing through handover between neighboring BSs is employed. This, however, has implications on load balancing among BSs.

In contrast, queue-aware relaying schemes offer the opportunity to circumvent the problem of large pathloss, due to heavy shadowing for instance, through in-cell routing. This is done by selecting the appropriate RS(s) and resource units for the associated hops. The authors in [38], used the aforementioned metric in a relay-enhanced single carrier CDMA network. In the next chapter, joint routing and fair scheduling algorithms employing a similar metric are presented. The algorithms represent different optimization formulations in an OFDMA-based relay-enhanced network [57], [90].

2.9 Channel Models and Numerical Examples

Radio propagation models have significant impacts on the performance of algorithms designed for wireless communication systems. To test new algorithms, the channel models need to be versatile to adequately represent the real-life environments in which the systems will be operating. For this reason, many research forums have considered channel models as an important part of their research efforts. The Wireless World Initiative New Radio (WINNER) project [92] invested a significant amount of time in developing spatial channel models that are an evolution of 3GPP's channel model. To capture a wide range of applications, the WINNER channel model extends from short range to wide area which covers indoor, typical urban micro-cell, typical urban macro-cell, sub-urban macro-cell, rural macro-cell and stationary feeder link and in the frequency range 2 to 6 GHz.

Frequency diversity and time selectivity are among the features that an RRM scheme could exploit to enhance the performance of the wireless network using OFDM signaling. Generating channel samples through IFFT/FFT [91] present a peculiar challenge and it is worth mentioning. The IFFT/FFT approach requires all channel samples to be generated and stored prior to simulating the allocation process. Thus, in a large number of subcarriers or subchannels, and potentially a large number of access and interference links (and/or due to the many fixed and nomadic relays), generating a large number of independent channel realizations can be computationally exhaustive³. We demonstrate the realization of versatile and realistic channel models that incorporate Rayleigh (Fig. 2.9) and Ricean (Fig. 2.10) fading correlated in time and frequency. Fig. 2.9 shows some realizations of time-frequency correlated Rayleigh fading channel and a 6-tap power delay profile. A channel model representing a low Doppler spread and line-of-sight feeder scenario is shown in Figure 2.10. In these

³Refer to Appendix A for simulation tips on how to generate large numbers of independent time-frequency correlated fading realizations.

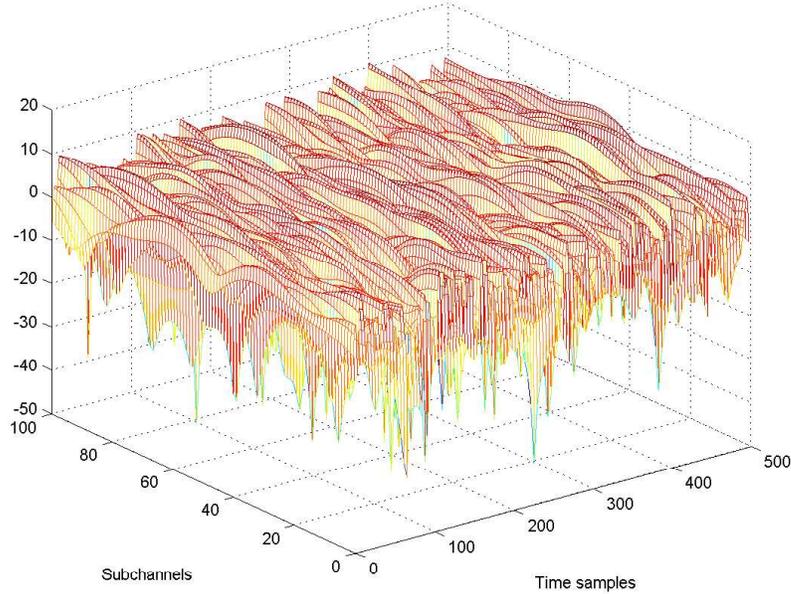


Figure 2.9: Time-frequency correlated Rayleigh fading realization for a speed of $v = 10$ Km/hr $f_c = 2.5$ GHz.

figures, the relative speed v (m/s) between the transmitter and receiver, the operating carrier frequency f_c (Hz) are given. Therefore, the max Doppler spread f_d (Hz) can be calculated using [91]

$$f_d = \frac{v}{c} f_c, \quad (2.8)$$

where c (in m/s) is the speed of light. The delay spread for the simulated channels can also be obtained from the power delay profile, $\mathbf{h} = [0 \ -1 \ -9 \ -10 \ -15 \ -20]$ dB and tap delays, $\boldsymbol{\tau} = [0.000 \ 0.300 \ 0.700 \ 1.100 \ 1.700 \ 2.500]$ μs . The essential features of the channel that can be exploited by OFDM technique is obvious in the two figures. The dynamic fading across frequency and time provides the opportunity to use the channel when it is most favorable for transmission. Despite this opportunity, some of the works in the literature rely only on the frequency diversity, and assume time-invariant channels with little or no mobility [8, 46]. However, in the single-cell relay networks in [43] and [50], time-frequency-correlated Rayleigh fading channels, with exponential power-delay profile for user links using Jakes' model, are used.

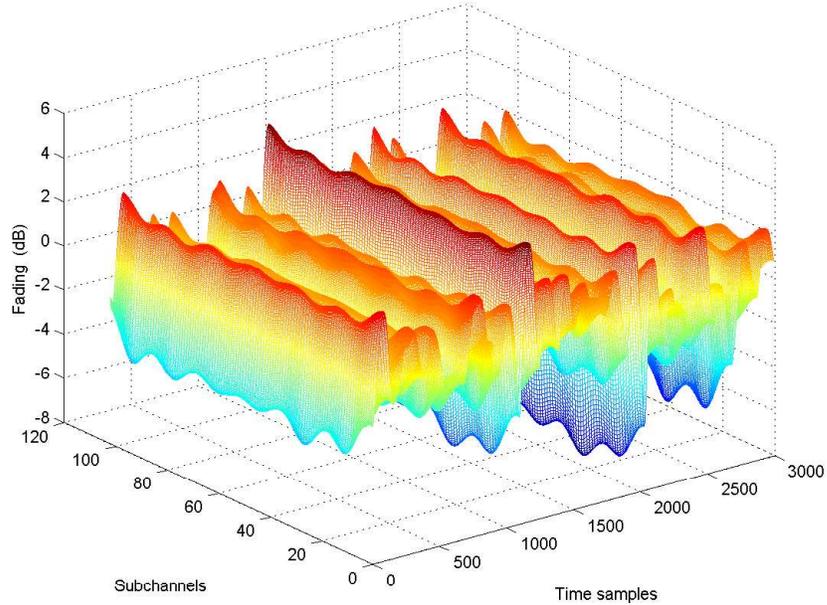


Figure 2.10: Time-frequency correlated Rician (K-factor of 10 dB) fading realization 144 subchannels $f_d = 4$ Hz.

2.10 Conclusions

This chapter provides a comprehensive survey of radio resource management schemes in relay-enhanced OFDMA-based wireless networks. Numerous publications have highlighted the crucial need for devising intelligent RRM schemes to harness the opportunities in future OFDMA relay networks where conventional schemes are not applicable. We addressed some of the opportunities, challenges, and terms associated with the migration from conventional cellular architecture to relay-enhanced. In general, the algorithms proposed in literature aim at exploiting the variations in wireless channels by adaptively allocating scarce communication resources to network entities

to either maximize or minimize certain network metric for some given constraints. Users' expectations are much higher given the advertised objectives and fairness obligation is stronger in such networks. We therefore discuss some fairness implementation techniques along with some example fairness metrics towards the design and performance evaluation of prospective fair radio resource allocation algorithms.

Some of the proposed algorithms are designed to reduce the required signalling overhead compared to the optimal solution, when the optimal solutions involve prohibitive complexity. For instance, there are on-going research efforts towards finding more efficient centralized RRM algorithms, since most of those proposed in the existing literature are overly complex. Furthermore, significant savings in overhead and system complexity can be obtained through distributed resource allocation schemes. Therefore, research in distributed RRM algorithms in OFDMA relay networks has started drawing attention.

Chapter 3

Fairness-aware RRM in Downlink OFDMA Cellular Relay Networks

3.1 Introduction

Through intelligent RRM, the combination of relaying and OFDMA techniques has the potential to provide high data rate to WSs or user terminals (UTs) everywhere, anytime. In contrast, conventional opportunistic schedulers will rarely serve UTs with bad channel conditions such as cell edge UTs; this defeats the notion of ubiquitous coverage targeted in future networks, and exposes the importance of fair RRM algorithms to facilitate location-independent service, especially when users subscribed to the same service class are charged similarly regardless of their channel conditions [31].

The opportunities in relaying and OFDMA techniques bring some interesting challenges due to the increased dynamics, degrees of freedom, resource reuse, and complexities incurred in resource allocation and interference management, especially in networks with large numbers of users and relays [11]. This fact highlights the importance of dynamic and intelligent radio resource management (RRM) schemes with efficient spectrum utilization [11], [58]. As such, the literature on RRM in OFDMA-based cellular relay networks is steadily growing discussing various schemes in terms of objective (user-centric or network-centric), processing and feedback (ranging from

fully centralized to distributed), as well as scope (considering systems with single cell/single relay to multicellular/multiple relays) [10].

As discussed earlier in Chapter 2, the RRM algorithms in the literature of OFDMA conventional networks can not be directly applied to relay-enhanced networks since the problem is not just a scheduling problem. Rather, it is in principle, a joint routing and scheduling problem. However, the vast majority of RRM schemes presented in the literature decouple in-cell routing and resource allocation for simplicity. Thus, limiting the opportunities in spatial diversity and channel dynamism the scheme could exploit. In fact, pathloss-based and distance-based relay selection are common and simple strategies, e.g., [37] and [48]. In addition, the desired user fairness may not be attained through the fairness-oriented schemes that rely solely on achievable and allocated capacities (channel-awareness only), e.g., proportional fair scheduling (PFS) [77], [93]. We have also observed in the literature the tendency to maximize the total cell capacity, e.g., [8] and [49] whereas capacity does not map directly to throughput due to the burst traffic. As such, imposing constraints to allocate fair shares of the cell capacity might not result in actual throughput fairness. The lack of traffic- or queue-awareness also prevents such schemes from exploiting the traffic diversity (statistical multiplexing) and accounting for previously relayed data that need to be rescheduled due to a practical ARQ protocol. Finally, the relay-based RRM algorithms developed for single-cell system models along with their performance results are not applicable to practical multicellular scenarios since inter-cell interference is not considered, e.g., [43], [44] and [46].

Although distributed schemes are more attractive, it is essential to seek outstanding performance benchmarks to which various decentralized schemes can be compared. Nevertheless, an interesting question arises; can we design centralized RRM algorithms that incur low complexity and save substantially in feedback overhead?

3.2 Contributions and Relevant Work

Devising dynamic traffic- or queue-aware RRM schemes tackling the joint routing and scheduling problem constitutes therefore a worthwhile yet challenging research opportunity. In [56], Tassiulas et al. laid a foundational theory on throughput-optimal scheduling in wireless multihop mesh networks incorporating queue-awareness into the scheduling policy which allocates resources dynamically to multicommodity flows. They showed that maximizing the sum of a queue length-based drift metric over all node pairs is the maximum throughput policy which stabilizes all network queues under the largest set of mean exogenous arrival rates for which the network queues can be stabilized. Nevertheless, the authors stressed that devising efficient algorithms to solve the optimization problem given the constraint set imposed by the system model of each particular application is important for implementation.

Several works have adopted throughput-optimal scheduling thereafter proposing scheduling policies for adhoc networks, non-OFDMA, or conventional (non-relaying) cellular networks with different optimization formulations. For instance, in [88] and [89], conventional cellular SDMA/TDMA and OFDMA networks are respectively considered thus eliminating the joint routing and scheduling aspect of such policies and limiting the queue stabilizing opportunities to the resource allocation at the BS.

While fairness is crucial to realize the desired service ubiquity and reliability in cellular networks, it should be noted though that throughput-optimal policies are not fairness-oriented in principle, as they aim at stabilizing all user queues under any heterogenous traffic flows within the system's capacity region. Therefore, in [94], a congestion control mechanism is proposed for a conventional cellular network to introduce user fairness through traffic policing, if the arrival rates at the BS are elastic (adaptive).

In [55], Neely et al. proposed a centralized dynamic routing and power control

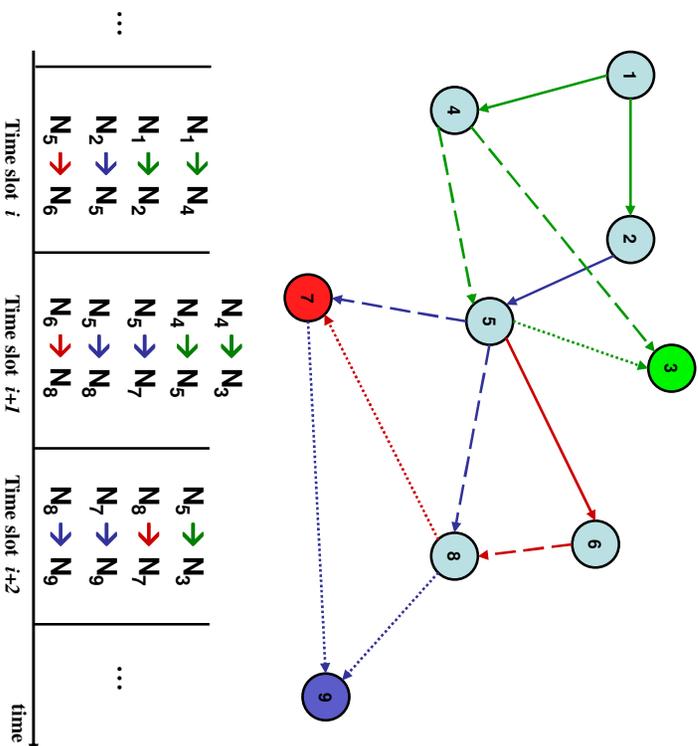


Figure 3.1: An example multihop packet radio network with multicommodity flows (shown in red, green and blue along with their respective sinks). The slotted time axis shows the traditional quasi-FDR protocol in throughput-optimal scheduling at the nodes N_2 , N_5 , and N_7 .

policy (DRPC) in a single-carrier adhoc network with multicommodity flows (see Fig. 3.1), rate adaptation, and node power budgets. In each time slot, the DRPC policy solves a *one-shot* optimization to allocate power to a set of links carrying the chosen commodities such that the sum metric is maximized. The authors did not however suggest ways to solve such an optimization under the node power constraints and the co-channel interference governing the achievable rates of these links. Therefore, without considering the power control dimension, a centralized joint routing and scheduling algorithm is proposed in [38] for the downlink (DL) of a single-carrier CDMA *cellular* relay network under *symmetric traffic arrival processes*. The authors suppose that throughput-optimal scheduling is a *fair* policy in such a case. It is assumed however that a route to the UT may comprise an indefinite number of hops. The algorithm also incurs high complexity and is not applicable to multi-carrier systems.

More importantly, the one-shot optimizations in [55], and similar works such as [88], [38] and [95], have no mechanisms to prevent outstanding queues -with relatively high mean arrival rates or with the same mean arrival rates yet experiencing high instantaneous bursts- from unnecessarily acquiring most, if not all, of the system resources during the subject time slot. That is because the backlog weights will not be affected by how many resources are allocated until the next time slot. Thus, in such scenarios resources are wasted while low traffic flows experience high latency until their backlog weights dominate; that indeed implies a limitation on the system's capacity (within which the policy is throughput-optimal) due to implementation. An enhanced DRPC policy is suggested in [55] where priorities of low traffic queues are enforced through some empirical scaling parameters at each node.

It is worth noting that *in the literature on throughput-optimal scheduling in multi-hop mesh networks, the commonly adopted (relaying) protocol assumes that the nodes are capable of transmitting and receiving different data concurrently on different channels*, e.g., node N_2 in slot i , N_5 in slot $i + 1$, and N_7 in slot $i + 2$, in Fig. 3.1. We refer to that traditional protocol as **the quasi full-duplex relaying (quasi-FDR) protocol** which is different from the theoretical full-duplex mode considered earlier in the literature [96], [97] . Hence, given their numerous desired features, it is quite interesting to study first the performance of traditional throughput-optimal policies (employing the quasi-FDR) when formulated to operate in an OFDMA cellular relay setup.

Therefore, in this chapter, we study a novel throughput-optimal formulation with a novel low-complexity centralized algorithm that achieve a ubiquitous coverage, high degree of user fairness and enables intra-cell load balancing under symmetric user traffic in the downlink of OFDMA cellular fixed-relay networks. The policy performs joint in-cell routing and scheduling *using only two-hop* relaying and prevents resource waste, in contrast to prior art, through efficient bit-loading constraints or iterative

optimization in the low-complexity algorithm. The scheme utilizes the opportunities provided in channel dynamism, spatial, and queue and traffic diversities. We also show how the feedback overhead is substantially reduced as compared to existing centralized schemes.

The journal paper [57], the patent fillings [98], the conference paper [99], as well as the technical report [100] are the outcomes of the work presented in this chapter. Open research directions and possible extensions of this work are provided in Chapter 7.

3.3 System Description and Assumptions

In the multicellular network, the BS serves K UTs either directly or through M RSs in a cell. All resources are available in each cell resulting in aggressive resource reuse. The total bandwidth is divided into N subchannels, each composed of a set of adjacent OFDM data subcarriers¹. The serving BS and each of the M RSs in a cell are equipped with K user-buffers. User packets arrive at the corresponding BS buffer according to the traffic model. The channel fading is assumed to be time-invariant within a frame duration. We first consider a generic scenario that is not restricted to a specific geographical deployment of RSs. Thus, potentially, any UT can be connected to any combination of the M RSs yet in only two hops as RSs are not allowed to exchange user data. Such unconstrained relay selection or ‘open routing’ exposes the ability of our routing strategy to dynamically settle for the best route(s) for each UT given an arbitrary relay deployment. We also present a constrained mode of operation for the routing strategy where geographical relay deployment can be exploited offering substantial savings in feedback overhead.

In the proposed scheme, a UT can receive from a group of nodes (BS and/or RSs),

¹The number of OFDMA subcarriers that comprise a subchannel is less than the expected coherence bandwidth of the channel.

and any node can transmit as well to multiple destinations, simultaneously, on different orthogonal subchannels. In line with the quasi-FDR protocol, any RS is assumed to have the ability to receive and transmit concurrently on orthogonal subchannels. A practical concern might arise if orthogonal transmit and receive subchannels happen to be close in frequency band. Since RSs are fixed, they can be deployed with two antennas (if necessary, at different elevations); a directional antenna for the feeder link from the BS and an omni-directional antenna to the UTs, thus, alleviating such concern.

Load balancing in relay networks, as discussed earlier in Section 2.8, is often associated with distributing the load evenly among the cell nodes. Thus, the number of OFDM subcarriers handled by a node is often employed in literature as a good estimate of the traffic load at that node [9], [44]. The following section describes the proposed scheme in detail and explains how the load balancing function is integrated.

3.4 The BS's Joint Routing and Fair Scheduling

The objective is to maximize the total cell throughput while maintaining throughput fairness among users. The idea is to operate a throughput-optimal scheduling policy, that stabilizes user queues at all nodes, in a system that receives equal inelastic mean arrival rates at only one source node in the cell which is the BS. Therefore, the fair behavior of such policy is a special case due to our cellular network system model where we consider that all users belong to the same service class and thus have the same mean arrival rates and the same QoS requirements. Such policy is perceived fair given a similar scenario in [38]. Since in principle throughput-optimal policies perform joint routing and scheduling of traffic dynamically without knowledge of the channel and traffic statistics, a maximization of the sum of the drift metric with proper constraints on frame-by-frame basis achieves our throughput and fairness objectives

and exploits the degrees of freedom in multiuser, spatial, and traffic diversities.

Let us define the drift (or loosely ‘demand’) metric for any node-UT link on subchannel n as the product of the achievable rate on that link and the queue length of the user’s buffer at that node, as follows

$$D_{n,m \rightarrow UT_k} = R_{m,k,n} Q_k^m, \quad m = 0, 1, \dots, M, \quad (3.1)$$

whereas the demand of any BS-RS link on subchannel n incorporates the queues at the BS (node 0) and those at the RS and can be expressed as

$$D_{n,BS \rightarrow RS_m} = R_{0,m,n} \max_k \{(Q_k^0 - Q_k^m)^+\}. \quad (3.2)$$

The function $(\cdot)^+$ sets negative arguments to zero. Q_k^m is the queue length of UT k at node m in bits or bytes (shown in blue bars in Fig. 3.2). Whereas $R_{m,k,n}$ and $R_{0,m,n}$ are the achievable rates on the links node $_m$ -UT $_k$ and BS-RS $_m$, respectively, on subchannel n . These rates are calculated, without loss of generality, using the continuous rate formula for adaptive modulation given as

$$R_{i,j,n} = W \log_2 \left(1 + \frac{-1.5 \gamma_{i,j,n}}{\ln(5 P_e)} \right), \quad (3.3)$$

where $\gamma_{i,j,n}$ is the the received SINR from source i at destination j on subchannel n . P_e and W are the target bit error rate and the OFDM subchannel bandwidth, respectively. As an alternative, either Shannon capacity formula (possibly with some practical SINR gap or penalty) or a discrete AMC lookup table can be used.

It can be easily shown that any monotonically increasing function of the metric, in its composite form, will result in the same RRA.

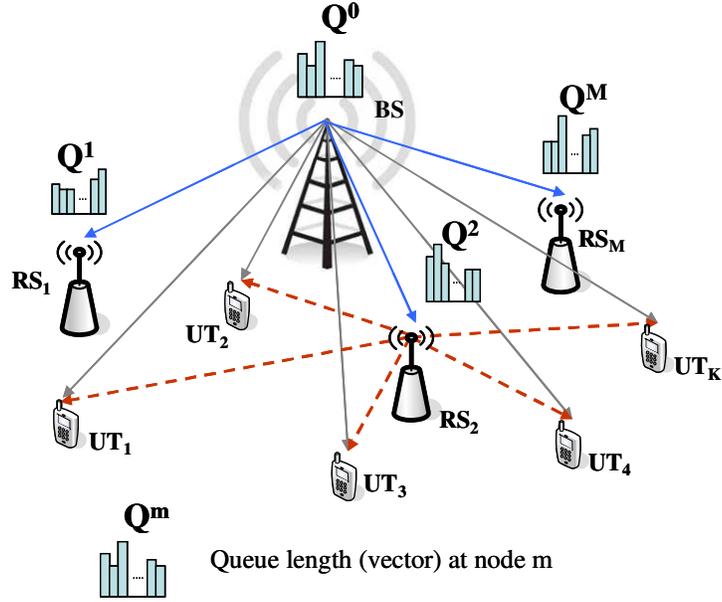


Figure 3.2: Example partial network showing the potential links of the BS and RS_2 on subchannel n .

3.4.1 Mathematical Formulation of the RRA at the BS

In order to maximize the total cell throughput while stabilizing user queues at all nodes, the RRA scheme needs to assign the subchannels with the highest capacities at any node to the outstanding queues at that node. This can be achieved by optimizing the assignment of subchannels to all links and the assignment of user buffers to feeder links so that the sum-demand is maximized at each allocation instant. The resource allocation at the BS can be formulated as a binary integer linear programming (BILP) problem as

$$\begin{aligned} & \max_{\rho, \gamma} \left\{ \sum_{n=1}^N \sum_{m=0}^M \sum_{k=1}^K \rho_{m,k,n} R_{m,k,n} Q_k^m \right. \\ & \left. + \sum_{n=1}^N \sum_{m=1}^M \gamma_{0,m,n} R_{0,m,n} \max_k \{ (Q_k^0 - Q_k^m)^+ \} \right\}, \end{aligned} \quad (3.4)$$

subject to the constraints

$$\rho_{m,k,n} \in \{0, 1\} \forall (m, k, n), \quad \gamma_{0,m,n} \in \{0, 1\} \forall (m, n), \quad (3.5)$$

$$\sum_{m=0}^M \sum_{k=1}^K \rho_{m,k,n} + \sum_{m=1}^M \gamma_{0,m,n} \leq 1 \quad \forall n, \quad (3.6)$$

$$\begin{aligned} \sum_{n=1}^N \sum_{k=1}^K \rho_{0,k,n} + \sum_{n=1}^N \sum_{m=1}^M \gamma_{0,m,n} &\geq \mu, \\ \sum_{n=1}^N \sum_{k=1}^K \rho_{m,k,n} &\geq \mu \quad \forall m \neq 0, m \in \mathcal{M}_{act}, \end{aligned} \quad (3.7)$$

$$\begin{aligned} T \sum_{n=1}^N \left(\rho_{0,k,n} R_{0,k,n} + \sum_{m=1}^M \gamma_{0,m,n} R_{0,m,n} \omega_k^m \right) &\leq Q_k^0 \quad \forall k, \\ T \sum_{n=1}^N \rho_{m,k,n} R_{m,k,n} &\leq Q_k^m \quad \forall (m, k), m \neq 0. \end{aligned} \quad (3.8)$$

In the above, $\rho_{m,k,n}$ is the k^{th} UT binary assignment variable to the m^{th} node, $m = 0, 1, 2, \dots, M$, on the n^{th} subchannel ($m = 0$ corresponds to BS, and the rest correspond to relays). The variable $\gamma_{0,m,n}$ is the m^{th} relay binary assignment variable to the BS node on the n^{th} subchannel whereas T is the transmission time of the downlink frame and $\mu = \lfloor N/|\mathcal{M}_{act}| \rfloor$ is the minimum number of subchannels to be assigned to any active node (BS or RS) where $\mathcal{M}_{act} = \{m : \sum_k Q_k^m \neq 0\}$. The binary indicator ω_k^m is 1 if user k has the highest queue difference between the BS and RS_m , and 0 otherwise. The constraints in (3.5) force the optimization variables to binary values while the constraints in (3.6) ensure that at most one link is active per subchannel. The constraints in (3.7) guarantee even distribution of subchannels among all nodes and hence balances the load. Finally, the constraints in (3.8), unlike the majority of works in the literature, e.g., [44], [43], and [48], ensure efficient bit-loading and prevent resource waste if the total capacity of the links withdrawing from a particular buffer is greater than the queue length at that buffer. Therefore, solving the optimization problem in such a novel formulation, results in the joint routing and

fair scheduling, guarantees efficient use of resources, and balances the load among cell nodes. A discussion on the routing strategy will follow in the next subsection. The unique aspects of the problem formulation leading to the outstanding performance of the proposed scheme are summarized as follows:

- A single linear objective function is maximized towards achieving a remarkable combination of both high ubiquitous throughput and user fairness, under the system model considered.
- The formulation does not rely on or imply any kind of preset routes, user partitioning, or resource partitioning, which are known to be suboptimal simplifying techniques.
- Dynamic routing and scheduling are performed jointly using the ‘differential backlog’ represented by the queue-length difference between BS and RSs [55]; this is analogous to the hydrostatic pressure between fluid tanks connected with pipes of different capacities which are controlled by the on-off assignment variables while UTs represent the relevant sinks of individual user flows.
- Traffic diversity is exploited through incorporating the buffer states; this does not require knowledge of the arrival process statistics.
- Load balancing between relay nodes is achieved jointly as well and not by rearranging the optimal allocation [9], [44].

The computational complexity, however, of such three-dimensional BILP problem is non-polynomial in time and can be approximated to $\mathcal{O}(((M + 1)K)^N)$. As such, the complexity might reach prohibitive limits in a system with high density of UTs and RSs given the expected high number of subchannels. Therefore, in the next subsection, we propose a low-complexity iterative algorithm that virtually updates the buffer states between iterations while satisfying all of the aforementioned constraints.

3.4.2 A Low-complexity Iterative RRA Algorithm

The formulated problem can be viewed as a three-dimensional assignment problem in which even subchannel-to-node assignments are required. The Hungarian algorithm [101] is an efficient solver, with polynomial complexity, for similar two-dimensional assignment problems and thus has been used in different scheduling algorithms for non-relaying networks, e.g., [7], [69], and [102]. Before we discuss how the Hungarian algorithm is applied to our assignment problem, we highlight the following facts: 1) Applying the Hungarian algorithm to an N -by- $(M + 1)$ profit matrix results in the optimal one-to-one assignment. 2) If all the N jobs ($> M + 1$) are required to be assigned such that each worker (node) handles almost the same number of jobs (load balancing) while his assignments are interdependent, due to time (or buffer size) constraints, a close-to-optimal solution is attained by applying the Hungarian algorithm $N/(M + 1)$ times (iterations) while eliminating the assigned jobs (subchannels). Note that each iteration is solved optimally.

If we visualize the three-dimensional space of the assignment variables in our BILP formulation, let the horizontal plane be N -subchannels-by- $(M + 1)$ -nodes then the vertical dimension represents all the links that node m may select from to activate on subchannel n . Mathematically, projecting onto the horizontal plane, through maximizing along the vertical axis for each pair (n, m) , does not change the combinatorial problem, i.e., does not change the optimal solution since there is no interdependency between the disjoint sets of links at different (n, m) pairs.

Since the resultant two-dimensional demand matrix is generally rectangular, we resort to the iterative approach as interdependency may exist between multiple subchannel assignments to the same node due to the buffer size constraint. The demand values are thus updated based on the result of the previous iteration before conducting

the next one. In that case, each iteration is solved optimally as a one-to-one assignment through the vertical maximization followed by the Hungarian algorithm. As a result, the performance of the whole algorithm is close-to-optimal. The algorithm is composed of the following steps which are executed each allocation instant:

1. The demand metric of each RS_m on subchannel n is calculated as the maximum of K potential links as follows:

$$D_{n,m} = \max_k \{R_{m,k,n} Q_k^m\}, \quad m = 1, 2, \dots, M. \quad (3.9)$$

Thus, $D_{n,m}$ is the best proposal of RS_m to use subchannel n while the UT associated with that maximum is marked as the candidate receiver. The demand metric for the BS node is the maximum metric of $M + K$ potential links and is expressed as

$$D_{n,0} = \max_j \{D_{n,BS \rightarrow j}\}, \quad (3.10)$$

where $D_{n,BS \rightarrow j}$ is calculated using (3.1) and (4.2), and j denotes any of the potential destinations. Thus, $D_{n,0}$ is the best proposal of the BS to use subchannel n . The destination associated with that proposal is marked as the candidate receiver. Note that if the destination is a RS, the UT that achieved the highest queue difference on that link is marked as well.

2. After calculating the $(M + 1)$ demand metrics on each subchannel, the algorithm solves a one-to-one optimization problem to maximize the total demand by applying the Hungarian algorithm to the $N \times (M + 1)$ demand matrix $[D_{n,m}]$ (See Fig. 3.3).
3. The algorithm virtually updates the affected UTs' queues according to the decisions of the previous iteration.

$$Q_k^{m^{(\iota+1)}} = (Q_k^{m^{(\iota)}} - \lfloor R_m^{(\iota)} T \rfloor)^+. \quad (3.11)$$

In the above, $Q_k^{m^{(\iota)}}$ is the input queue length to iteration ι and $R_m^{(\iota)}$ is the rate of

the link assigned by the Hungarian algorithm to node m as a result of iteration i . Note that the queues at destination RSs are not incremented between iterations because the transmissions on all subchannels occur simultaneously and the algorithm has to obey the causality law.

4. The rows corresponding to the assigned subchannels are eliminated.
5. Steps 1-4 are repeated for the unassigned subchannels until all enqueued packets are scheduled or the subchannels are exhausted.

Due to the one-to-one assignment, each iteration will only assign $M + 1$ subchannels to the $M + 1$ active nodes². As a result, each node is linked to only one destination per iteration; this prevents, along with step 3, the same queue length from being involved in the activation of more than one link as discussed earlier. Furthermore, if $N \bmod (M + 1) = 0$, each node will be assigned exactly $N/(M + 1)$ subchannels. Hence, load balancing is inherent in the algorithm. The discussion on simulation results will highlight the benefits of the load balancing feature.

For further illustration, a pseudo code for the iterative algorithm follows where \mathcal{U} , \mathcal{N} , \mathcal{K} , and \mathcal{M} denote the sets of unassigned subchannels, all subchannels, UTs, and RSs, respectively.

Routing of user data is thus performed dynamically and jointly with its resource allocation. Such dynamic routing strategy uses the maximum differential backlog represented by $\max_k \{Q_k^0 - Q_k^m\}$ to establish the routes. Several works have employed this dynamic routing strategy such as [38], [55], and [94], based on the throughput-optimal link scheduling policy developed in [56] for multihop packet radio networks where routes can comprise indefinite number of hops. However, by ‘open routing’ in the cellular network we mean that a UT can be connected to any set of RSs while the algorithm is not informed *a priori* of which RS(s) to use for that UT. Note that

²The number of active nodes is generally $|\mathcal{M}_{act}|$ which is usually equal to $M + 1$ in a loaded network under open routing.

Pseudo-code for the quasi-FDR RRA algorithm at the BS

```

Initialization:  $\mathcal{U} = \mathcal{N}$ 
while  $\|\mathcal{U}\| \neq 0$  and  $\sum \mathbf{Q}^m \neq \mathbf{0}$ 
  for each  $n \in \mathcal{U}$ 
    for  $m = 1$  to  $M$ 
       $D_{n,m} = \max_k \{R_{m,k,n} Q_k^m\}$ 
       $\kappa_{n,m} = \arg \max_k \{R_{m,k,n} Q_k^m\}$ 
       $D_{n,0 \rightarrow m} = R_{0,m,n} \max_k \{(Q_k^0 - Q_k^m)^+\}$ 
       $\kappa^m = \arg \max_k \{Q_k^0 - Q_k^m\}$ 
    end for
     $D_{n,0 \rightarrow k} = R_{0,k,n} Q_k^0$ 
     $D_{n,0} = \max_j \{D_{n,0 \rightarrow j}\}, \quad j \in \mathcal{K} \cup \mathcal{M}$ 
     $\kappa_{n,0} = \arg \max_j \{D_{n,0 \rightarrow j}\}$ 
  end for
  %  $\mathbf{D} = [D_{n,m}]$  is the demand matrix.
   $(\hat{\mathbf{n}}, \hat{\mathbf{m}}) \leftarrow \mathbf{Hungarian}(\mathbf{D})$  % Vectors of indices
   $\mathcal{U} = \mathcal{U} - \{\hat{\mathbf{n}}\}, N_{assigned} = \|\hat{\mathbf{n}}\| = \|\hat{\mathbf{m}}\|$ 
  %  $N_{assigned} \leq \min\{M + 1, \|\mathcal{U}\|\}$ 
  for  $i = 1$  to  $N_{assigned}$ 
     $\hat{n} = \hat{\mathbf{n}}(i), \hat{m} = \hat{\mathbf{m}}(i), \hat{r} = \kappa_{\hat{n}, \hat{m}}$ 
    if  $\hat{r} \in \mathcal{M}$  then
       $\hat{k} = \kappa^{\hat{r}},$ 
       $Q_{\hat{k}}^0 = (Q_{\hat{k}}^0 - [R_{0, \hat{k}, \hat{n}} T])^+$ 
    else
       $Q_{\hat{k}}^{\hat{m}} = (Q_{\hat{k}}^{\hat{m}} - [R_{\hat{m}, \hat{k}, \hat{n}} T])^+$ 
    end if
  end for
end while

```

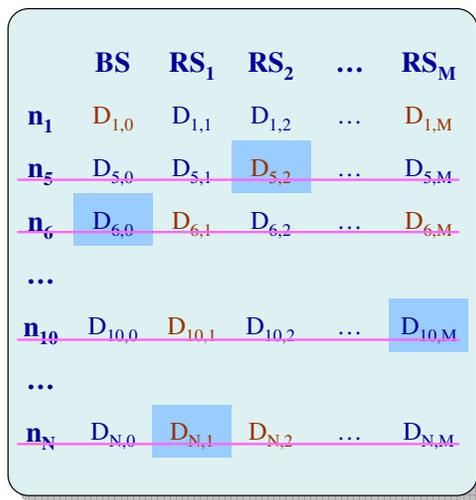


Figure 3.3: The demand matrix during one iteration. Rows with assigned entries are crossed out and eliminated. Red entries reflect on the queue updates due to the previous iteration.

a route is comprised of two hops only as RSs are not allowed to exchange user data amongst them. Therefore, in the open routing mode, initial accumulation of the user’s data may occur at some RS(s). For instance, let us assume that RS_M in Fig. 3.2 has a heavily shadowed link to UT₃ while the BS has forwarded some UT₃ data to RS_M, In this situation, these packets neither will be forwarded to UT₃ nor will they be absorbed by a neighboring RS. However, the algorithm exploits the presence of these trapped data, as they reflect on the quality of the second-hop link, by reducing the likelihood of further nominating UT₃ data on BS-RS_M feeder link, irrespective of the channel. That is, while some other user queues at RS_M are being discharged from one iteration to another, UT₃ may no longer achieve the maximum difference $Q_k^0 - Q_k^M$. The algorithm therefore, possesses the ability to learn from the previous forwarding mistakes to improperly selected RSs in previous iterations.

To further demonstrate the feasibility of the dynamic routing and its learning ability, we examine another mode of operation for the dynamic routing strategy named ‘constrained routing’ and compare it to the open routing mode. In that mode, routing constraints are imposed on BS-RS transmissions accounting for the geographical

distribution of the RSs and user locations. As such, the dynamic routing is allowed to operate on only $M_{cnst} (\leq M)$ closest RSs to each UT; this is done by ignoring the user buffers at the irrelevant (far) RSs while calculating the differential backlogs³. Intuitively, faster routing convergence is expected due to fewer forwarding mistakes. More interestingly, the improvement comes along with substantial savings in feedback overhead due to the eliminated links as discussed in Section 3.7.

The computational complexity has been significantly reduced using the iterative algorithm since each iteration optimizes a two-dimensional assignment problem using the Hungarian algorithm with polynomial complexity of $\mathcal{O}(N_u^3)$. This is because $[D_{n,m}]$ is rectangular with the number of unassigned subchannels N_u usually greater than the number of nodes ($M + 1$). Given that $M + 1$ rows (subchannels) are eliminated in each iteration, the maximum number of iterations required to assign all resources (if necessary) is $\lceil \frac{N}{M+1} \rceil$. Therefore, the complexity of iterations rapidly decreases in cubic polynomial manner as N_u decreases.

Since the total complexity of step 1 is $\mathcal{O}(MK)$, the complexity of the whole algorithm is loosely upper-bounded by $\mathcal{O}(\frac{N^4}{M+1})$, which is polynomial with substantial complexity reduction as compared to the BILP problem. A more precise complexity estimate is down to $\mathcal{O}(\frac{N^2(N+M+1)^2}{4(M+1)})$; that approximation holds for reasonable K satisfying $MK \ll N^2$. Note that both are bounds because the algorithm may incur less complexity if it stops before allocating all subchannels. Unlike the majority of algorithms, the complexity decreases as M , the number of RSs, increases. For the limiting case $M + 1 = N$, both estimates coincide at the asymptote $\mathcal{O}(N^3)$ which implies the optimal one-shot Hungarian solution.

³This selection of relay sets is adopted for simplicity. A more appropriate selection could be based on pathloss rather than distance only.

3.5 PFS-based RRM in OFDMA Relay Networks

In this section we consider the Proportional Fair Scheduling (PFS) concept [77], [76]. PFS is known in literature to provide an efficient throughput-fairness tradeoff for conventional (non-relaying) networks. Therefore, integrating PFS with the most commonly adopted RRM techniques in the literature of OFDMA relay networks represents a reasonable reference scheme. The generic framework is to partition the UTs into clusters around the chosen serving nodes (BS and RSs), partition resources among nodes accordingly, then allow individual nodes to perform PFS as adopted in OFDMA-based systems [49], [80]. Some essential details of the PFS-based scheme are shown in Fig. 3.4 and summarized as follows:

- The closest serving node is chosen by the UT. As such, the direct BS transmission to K_0 UTs occurs within a radius of half the distance between the BS and RSs while $K_r = \sum_{m=1}^M K_m$ UTs are relayed.
- Based on such connections, the BS reserves $N_0 = N \frac{K}{K_0 + 2K_r}$ subchannels to allocate among its direct UTs and feeder links. The remaining $N - N_0$ subchannels are partitioned among the RSs in proportion to the numbers of their connected UTs. The total power available at each node is divided equally among the subchannels of its allocated partition.
- The PFS at each node updates the average rates after allocating a subband of the available subchannels. The number of the subbands S is a parameter in the implementation of PFS in OFDMA-based systems [78] [87], and affects the choice of the averaging window size T_p in (3.12). This is a low-complexity implementation of PFS in multicarrier systems [93].
- At each RS_m , subchannel n is assigned to user $k^* = \arg \max_k \frac{R_{m,k,n}}{\overline{R}_k(\iota-1)}$, $k \in \mathcal{K}_m^a$, where \mathcal{K}_m^a is the set of active UTs (with buffered data at RS_m), $\overline{R}_k(\iota)$ is the exponentially weighted average rate of user k after allocating the current

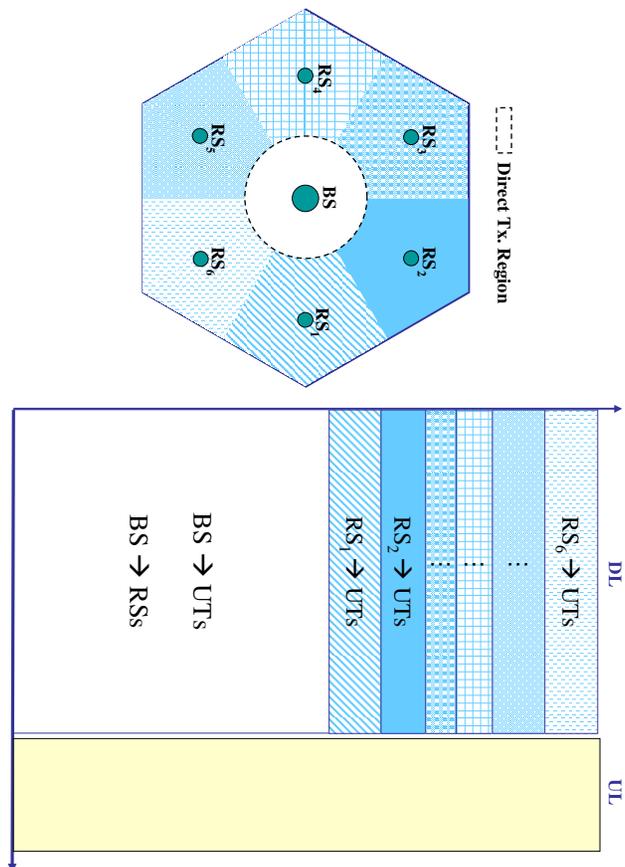


Figure 3.4: User and frequency partitioning in the reference relay-aided PFS scheme.

subband as defined in (3.12). The relevant user buffer is updated after the subchannel assignment as,

$$\overline{R}_k(i) = (1 - \frac{1}{T_p})\overline{R}_k(i-1) + \frac{1}{T_p} \sum_{n \in \mathcal{C}_k} R_{m,k,n}, \quad (3.12)$$

where \mathcal{C}_k is the set of the subchannels assigned to the UT $_k$.

- At the BS, subchannel n is assigned to one of the contending destinations (direct UTs and RSSs); $j^* = \arg \max_j \frac{R_{0,j,n}}{R_j(i-1)}$, $j \in \mathcal{K}_0^a \cup \mathcal{M}^a$, where \mathcal{M}^a is the set of active RSSs in which any RS has at least one connected UT with buffered data at the BS and $\overline{R}_j(i)$ is defined as

$$\overline{R}_m(i) = (1 - \frac{1}{T_p})\overline{R}_m(i-1) + \frac{1}{T_p |\mathcal{K}_{0,m}^a|} \sum_{n \in \mathcal{C}_m} R_{0,m,n}, \quad (3.13)$$

where \mathcal{C}_m is the set of the subchannels assigned to the feeder of RS $_m$. If the subchannel is assigned to the feeder of RS $_m$, the buffered data at the BS of some user $k \in \mathcal{K}_{0,m}^a \subseteq \mathcal{K}_m$ is scheduled on that feeder following a round-robin sequence. The relevant user buffer is updated after the subchannel assignment.

3.6 Simulated Network Performance

3.6.1 Simulation Models and Parameters

Matlab system-level simulations have been conducted. The Network and channel parameters are given in Table 3.1.⁴ The cellular network consists of 19 non-sectorized hexagonal cells enhanced with 3 or 6 RSs per cell. These relays are placed at a distance of 0.65 of the cell radius from the BS and with a uniform angular spacing. UTs are uniformly distributed within the cell area. Independent Poisson packet arrival processes are assumed at BS queues. The average arrival rate is 632 packets (188 bytes each) per second per UT. The path-loss model used is $PL = 38.4 + A \log_{10}(d)$ where $A = 23.5$ for BS-RS links and $A = 35.0$ for all other links. RSs transmit to UTs with an omni-directional antenna and receives with a highly directive antenna from the BS. Independent lognormal shadowing is assumed for all links but with different standard deviations. Time-frequency correlated Rician fading is assumed for (LOS) BS-RS links while all other (NLOS) links are assumed to experience time-frequency correlated Rayleigh fading.

3.6.2 Simulation Results and Discussion

Figure 3.5 shows scatter plots of UT time-average throughput against UT distance from the BS for 6 and 3 RSs with 25 UTs/cell. Each point in the plot represents the time-average throughput (over 100 allocation time frames) for a particular UT within a drop with fixed location and shadowing. The time average is calculated over the downlink frame duration which is 2/3 of the total TDD frame duration. Statistics are collected from 7 cells (the center cell and the surrounding 6 cells) for each of 30 drops. The performance of the proposed algorithm in its open and constrained

⁴Adopted from the WiMAX Forum based on IEEE 802.16e. The pathloss model, RS antenna pattern, and BS-RS channel power delay profile are adopted from the WINNER project: www.ist-winner.org.

Table 3.1: Simulation parameters

Parameter	Value
BS-BS distance	1 Km
RS distance from BS	$0.65 \times$ cell radius
UT min. close-in distance to BS	35 m
BS Tx. antenna gain	15 dB
RS Tx. antenna gain	10 dB
RS Rx. antenna $\theta_{3dB} = 20^\circ$	$\pi/9$
UT Rx. antenna gain	0 dB
Shadowing σ for NLOS links	8.9 dB
Shadowing σ , for LOS links (BS-RS)	4 dB
Rician K-factor for BS-RS links	10 dB
Carrier frequency	2.5 GHz
Total bandwidth	20 MHz
UT mobility	20 Km/hr
BS-RS links max. Doppler spread	4 Hz
Number of channel taps	6
Number of channel taps (BS-RS)	8
TDD frame length	2 msec
Downlink : Uplink ratio	2:1
DL Tx. time in OFDM data symbols	11 symbols
OFDM subcarrier bandwidth	10.9375 KHz
OFDM symbol duration	102.86 μ sec
Subchannel width	18 subcarriers
Total number of subchannels	102
CR-QAM target BER	10^{-3}
Noise power density at Rx. nodes	-174 dBm/Hz
BS total Tx. power P_B	46 dBm
RS total Tx. power P_R	37 dBm
PFS averaging window size T_p	5
PFS number of subbands S	7
PFS radius of direct Tx. region	$0.325 \times$ cell radius

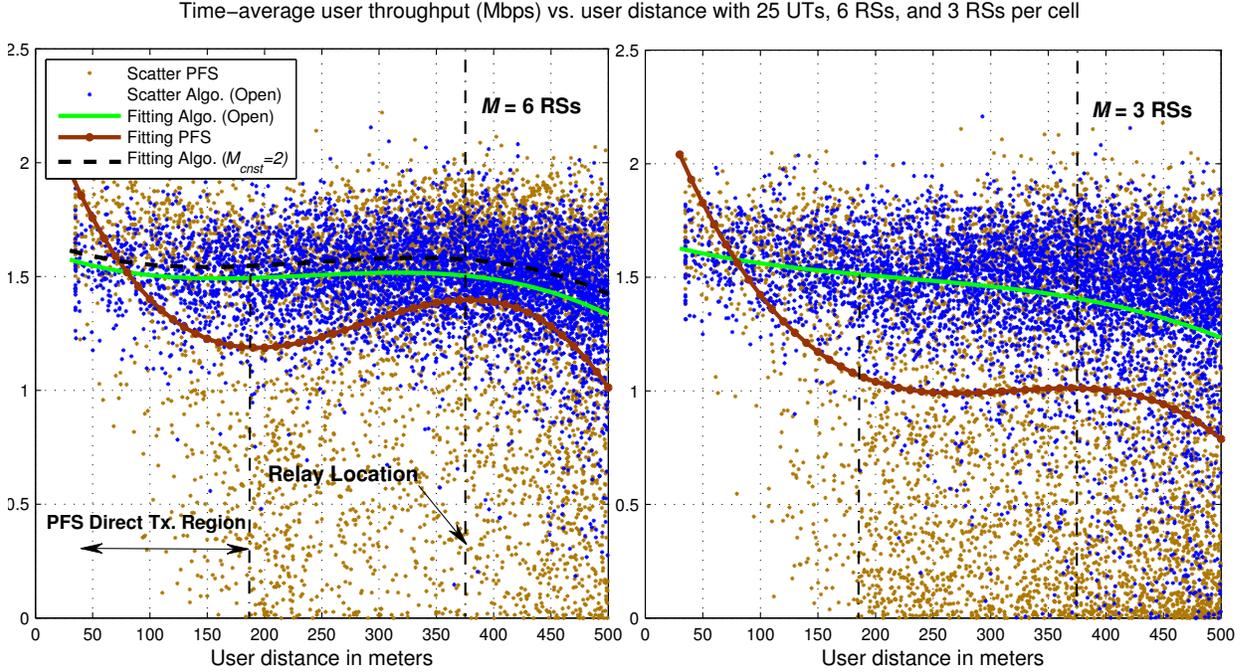


Figure 3.5: Time-average user throughput as function of user location and shadowing with 25 UTs/cell using 3 and 6 RSs.

routing modes and that of the reference PFS scheme are compared. The distance-based conditional mean of user throughput is approximated by the 3^{rd} -degree fitting curves of the scatter points as a means of averaging out shadowing. For the proposed scheme, the uniform average throughput across the cell area is clearly evident and demonstrated by the almost flat performance from BS to cell edge. This implies that a fair service and ubiquitous coverage are provided for all users regardless of their locations, channels, and interference conditions. Some throughput gain is further achieved when the algorithm operates with 6 RSs in its more practical constrained routing mode due to better routing convergence.

On the contrary, the coverage of the PFS reference scheme is significantly distance dependent as the mean throughput depreciates when users move away from the serving node, especially at the cell edge. That is mainly due the fact that spatial

diversity is not exploited (due to static routing) while scheduling a UT on the available subchannels partially exploits the frequency diversity and, moreover, may not overcome large pathloss (e.g., due to heavy shadowing) which dominates all the UT's subchannels. This results in a very poor time-average throughput for such UT (i.e., points at the bottom of the scatter plot). Whereas, the scatter points for the proposed algorithm have high throughput and narrow spread. This indicates the ability of the dynamic routing strategy to find the appropriate path(s) for such UTs and to deliver a fair service. The difference in performance is further emphasized in the scenario with 3 RSs as more users are expected to have poor link qualities from their serving RSs in the PFS scheme. However, the proposed scheme still offers a reasonable ubiquity and substantial throughput gains over the reference scheme, especially at the cell edge.

Note that the traditional PFS is not queue-aware and expected to provide performance inferior to the shown here where the PFS at a serving node excludes the users with empty buffers. Although such practical constraint enables the reference PFS to partially exploit the traffic diversity, it does not prevent resource underutilization. Figure 3.6 shows the CDF of the time-average throughput with both the open and constrained routing modes. Different M_{cnst} are considered. The 5th-percentile throughput of such CDFs is associated with cell-edge users in LTE evaluation methodology [80], [103]. First, a substantial 5th-percentile throughput advantage of 540% is realized for the proposed scheme in its open routing mode as compared to the PFS scheme which possesses a poor lower tail behavior. Second, the performance gaps between different cases of the constrained routing mode and the open mode of the dynamic routing strategy are generally narrow (8.6% at most). Such close performance implies that the open routing mode has an inherent capability of avoiding the poor routes to the UTs using the differential queue length information as discussed earlier in Subsection 3.4.2. However, some throughput losses are inevitable

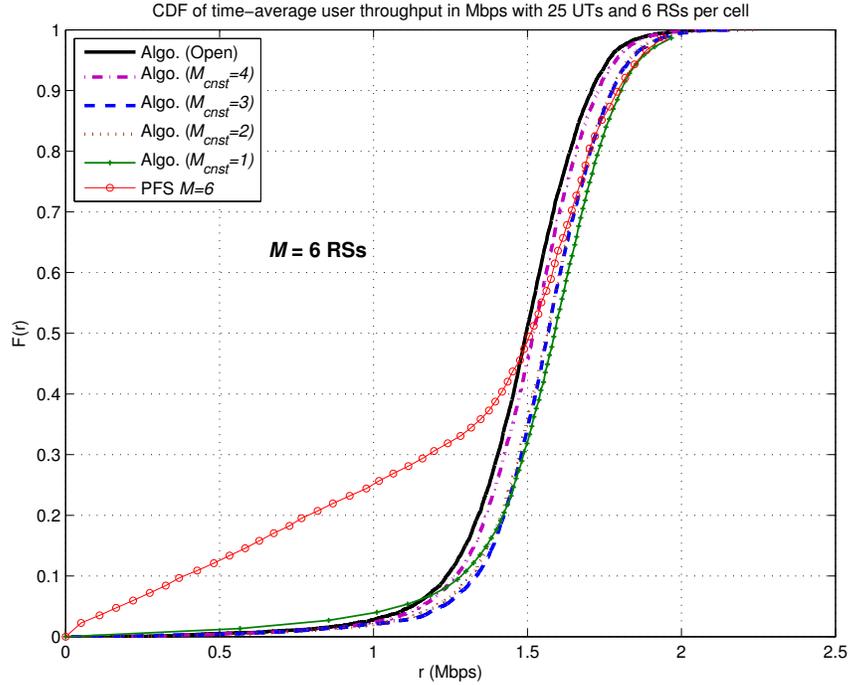


Figure 3.6: CDF of the time-average user throughput with 25 users and 6 RSs per cell.

due to initial forwarding mistakes and after occasional improvements of the poor links (due to small-scale fading and/or co-channel interference). Note that the constrained routing mode also utilizes the same learning ability to establish routes using fewer yet better candidate RSs for each UT. This can be observed in Fig. 3.7 (Fig. 3.6 with zoom-in).

In Fig. 3.7, a slight cell-edge throughput improvement of 2.76% is attained by excluding the farthest 2 RSs ($M_{cnst} = 4$) as compared to the open routing mode. Using only the 3 closest RSs ($M_{cnst} = 3$), yields the best improvement (8.6%) as the ‘far’ RSs with potentially poor links to UTs have been excluded along with the associated throughput loss. As expected, further elimination of RSs reduces the spatial diversity the dynamic routing exploits and thus the performance degrades slightly from $M_{cnst} = 3$ to $M_{cnst} = 2$ and significantly at $M_{cnst} = 1$, where only the closest RS is allowed, resulting in a degradation of 2.9% relative to the open mode.

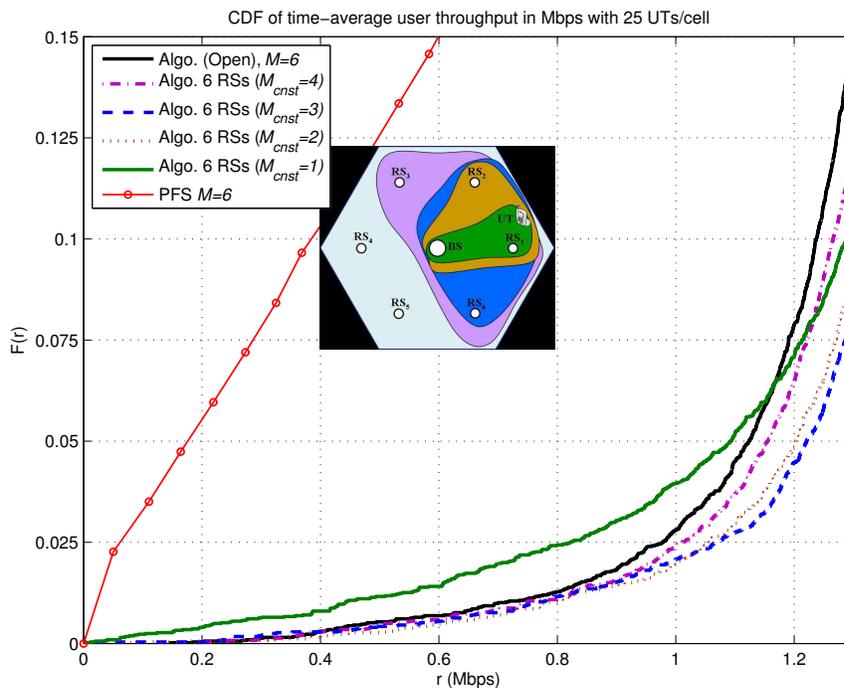


Figure 3.7: Lower tail behavior of the CDF of the time-average user throughput with 25 users and 6 RSs per cell.

As an alternative interpretation of these results; at a target average throughput of 1 Mbps, the outage probability of the proposed scheme ranges between 2.8% in the open mode and 2% in the constrained mode $M_{cnst} = 2$ or 3 as compared to 25.6% with the PFS scheme.

Figure 3.8 shows the total average cell throughput, as function of the number of UTs per cell, employing 3 and 6 RSs. The behavior in these curves is in agreement with the multiuser diversity concept and emphasizes the ability of the proposed scheme to maximize the total cell throughput by exploiting the multiuser, frequency, spatial, and traffic diversities. We employ the IEEE 802.16m fairness index [35] and Jain's fairness index [47] to assess the performance of the proposed and the reference schemes in terms of the time-average fairness and the long- and short-term fairness, respectively. Although the two metrics have different mathematical properties, both

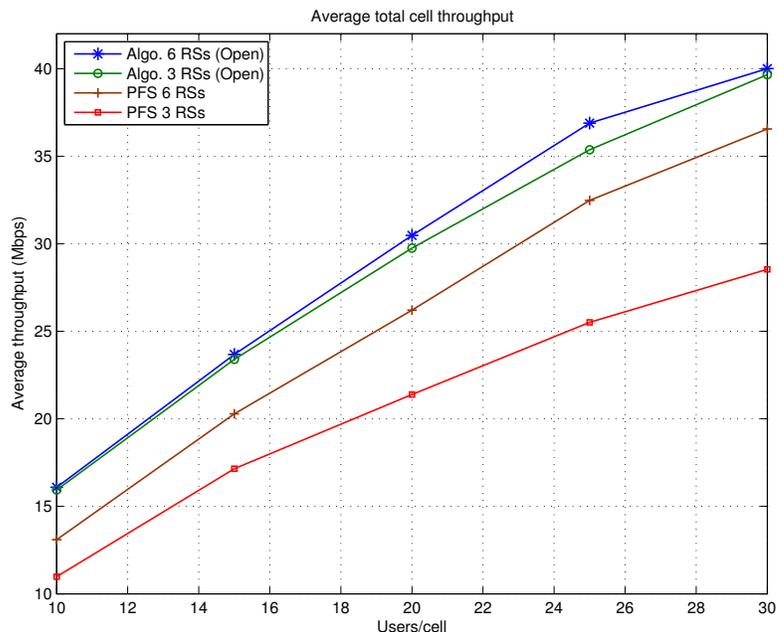


Figure 3.8: Total average cell throughput for the competent schemes.

metrics are symmetric for all user rates, and in both, ideal fair situation will be indicated by exact unity values. This is in line with our system model considering users with same priority. The time-average throughput rates, for 15 and 25 UTs with 3 and 6 RSs, are collected from all drops to plot the CDFs shown in Fig. 3.9. It can be observed that the same outstanding fairness behavior is achieved by the open and constrained routing modes with $M = 6$ at the different loading levels. While it becomes more difficult to maintain fairness as the number of users increases, there is insignificant degradation with the proposed scheme at 25 UTs/cell as opposed to the PFS scheme. Furthermore, reducing the number of RSs to 3 with 15 UTs has almost no impact on the fairness behavior of the proposed scheme and only a slight degradation with 25 UTs/cell; this however does not hold for the PFS scheme.

Jain’s index has been widely used in relevant works, e.g., [7], [80], and [104], and it is recommended by the WiMAX forum [105] for fairness assessment of proponents’ algorithms. It is here defined as

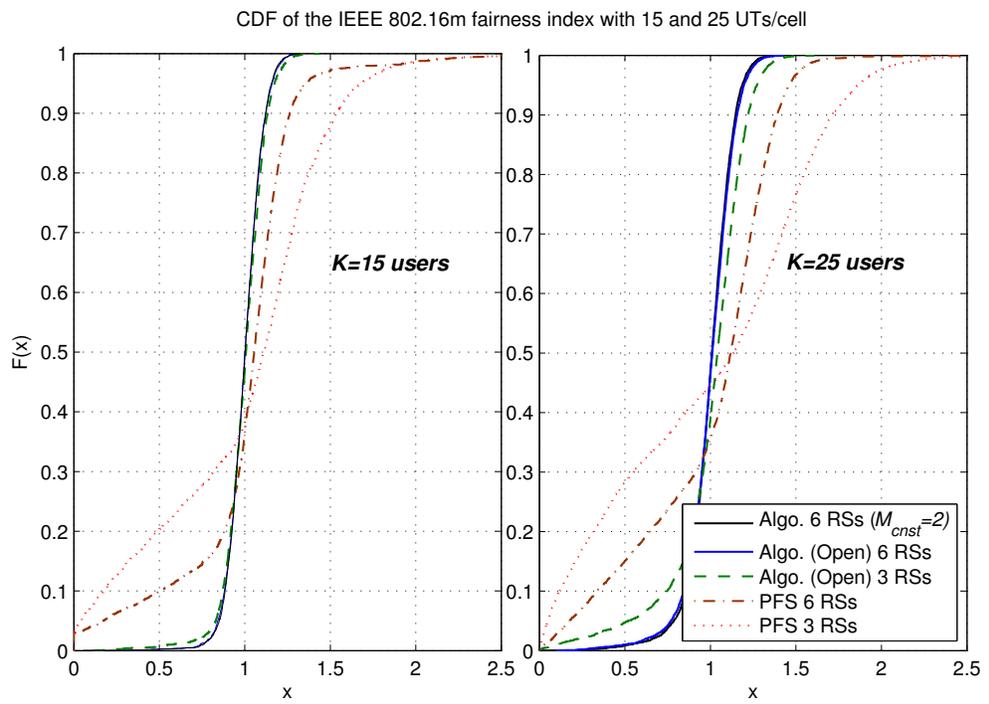


Figure 3.9: Time-average fairness using the IEEE 802.16m index with different numbers of users and RSs.

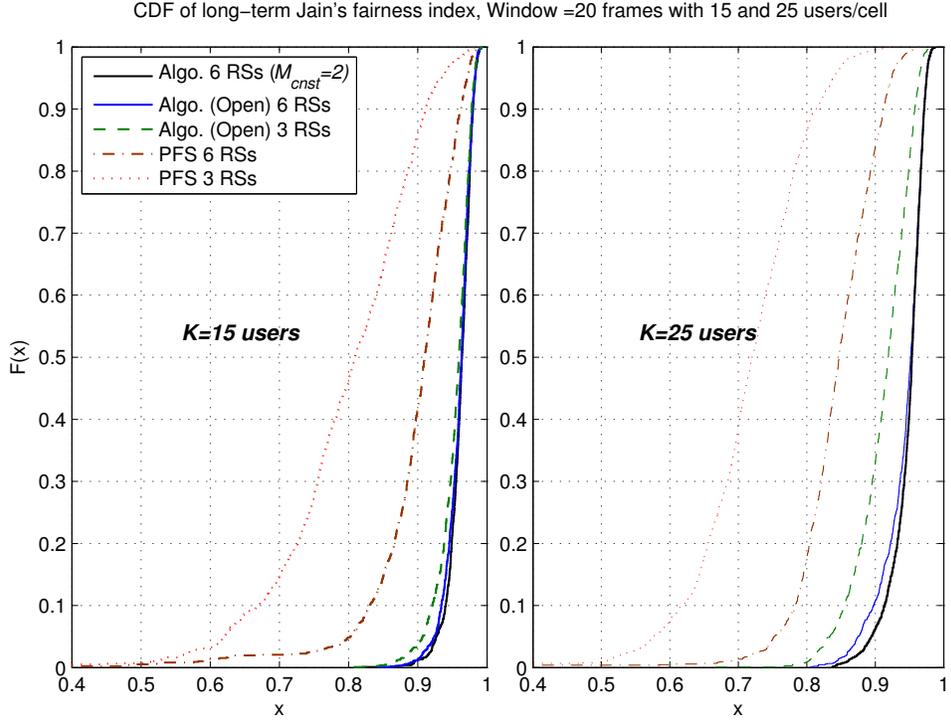


Figure 3.10: Long-term fairness using Jain's index with different numbers of users and RSs.

$$x_{w_j} = \frac{\left(\sum_{i=1}^K r_{i,w_j}\right)^2}{K \sum_{i=1}^K r_{i,w_j}^2}, \quad (3.14)$$

where r_{i,w_j} is the i^{th} user's average throughput rate achieved during the j^{th} time window w_j . As such, the index is a positive fraction that is lower-bounded by $1/K$. Therefore, in Fig. 3.10, the closer the CDF to a unit step at unity the more long-term fairness the scheme achieves after a time window of 20 frames. Although short- and long-term fairness are much more stringent than time-average fairness, the relative behavior observed in Fig. 3.9 matches that observed in Fig. 3.10 for long-term fairness.

Generally, queue-awareness allows RRM schemes to compensate the overlooked

user buffers, if any, and potentially improve user fairness, at least, in the long term sense. However, PFS relies on metrics based solely on allocated channel capacities. Furthermore, under static routing, as the number of users increases with fewer RSs employed, more UTs links to the serving nodes experience large pathlosses and thus only low achievable rates are left for the PFS to apply its fairness criterion. In contrast, the proposed scheme circumvents the problem of heavy shadowing and/or large pathloss through dynamic in-cell routing. This explains its ability to further improve fairness as time evolves and to exploit the spatial diversity when the number of RSs is increased, yielding such a wide gap in performance as compared to the reference PFS scheme.

It is worth mentioning that such outstanding performance in terms of throughput and fairness is achieved without overloading any node in the system. We demonstrate the intra-cell load-balancing behavior of the proposed scheme through Fig. 3.11 which shows a normalized histogram for the number of subchannels assigned to each node in a cell of a BS and 4 RSs during a drop of 100 time frames. It can be observed that each node is persistently assigned 20 subchannels, what is equal to $\frac{N}{M+1}$. We note that this is achieved with the exception that if a node has no buffered data, then this node is excluded from the assignment and the load balancing is maintained among the active nodes only. Thus, the very limited perturbations shown deliberately in the figure occur only during the first few initialization frames when RSs start with empty buffers while resources are assigned among the BS and only the RSs that have received data to forward. In addition to distributing the traffic load among cell nodes, the load balancing feature also spatially spreads (randomizes) the co-channel interference across the network exploiting the uniform geographical relay deployment.

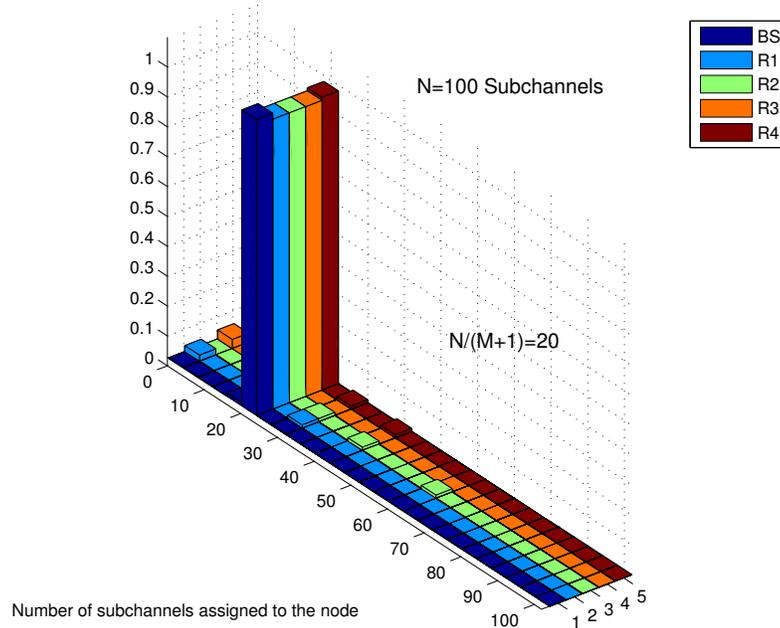


Figure 3.11: Normalized histogram of the number of subchannels assigned to transmitting nodes.

3.7 Signalling Overhead and Delays

In order for an RRM scheme to exploit the multiuser and frequency diversities, the allocation process should be conducted periodically with a period not greater than the shortest channel coherence time which is determined by the highest user mobility supported. Therefore, if the allocation is conducted at the beginning of each TDD frame, the feedback is required that frequent for maximum mobility users, e.g., 90 Km/hr based on our adopted frame duration as per the WiMAX Forum. However, for lower mobility, the feedback can be acquired as less frequently as each $\lfloor T_c/T_F \rfloor$ frames. That is the maximum integer number of TDD frames less than the user's coherence time of the channel (4 TDD frames in the simulated scenario). As such, the RRA algorithm can be invoked that often while the allocation result will be applied to the transmissions of the intermediate frames until the following allocation instant. Such relaxed resource allocation, however, less exploits the traffic diversity for highly burst traffic. We hereby discuss the following items to quantify the amount of feedback

information required each allocation instant:

- Implementing the constrained routing mode provides substantial savings in feedback overhead. That is because no feedback is required from the UT for the eliminated RS-UT links.
- In practice, AMC lookup tables are used and therefore reporting the indices of the achievable AMC levels per link significantly saves in signalling overhead as compared to reporting a wide range of continuous SINRs; this applies to both UTs and RSs.
- Our dynamic routing strategy, either in open or constrained mode, allows the UT to be connected to more than one node; having many users per cell, this implies that only very few subchannels are used for each node-UT link. As such, with potentially marginal performance losses, further savings in overhead can be achieved if UTs report only the ‘best’ fraction of subchannels in term of achievable rates⁵. Our studies considering only the best 50% of link subchannels show no performance degradation, even when the number of reported links is limited by constrained routing as shown in Fig. 3.12 and Fig. 3.13.

The following formula can be used to estimate the feedback overhead in the system taking into account the previous items:

$$T_{OH} = \frac{N_{CSI} (M_{cnst} + 1) n_{AMC}}{T_F [T_c/T_F]} \quad bps. \quad (3.15)$$

In the above, N_{CSI} and M_{cnst} denote the number of reported subchannels per link and the number of RSs allowed for the constrained routing, respectively. Whereas n_{AMC} denotes the number of bits used to indicate the index of the achievable AMC mode on a subchannel.

⁵Either a fixed number of the best subchannels or every subchannel whose quality is above a certain threshold.

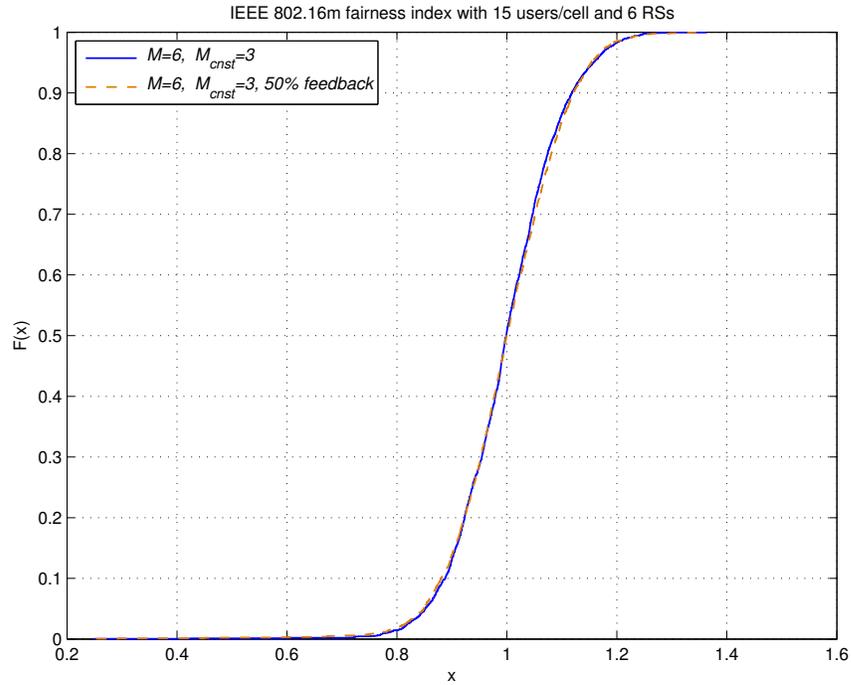


Figure 3.12: CDF of time-average user throughput with 15 users/cell and 6 RSs showing no performance degradation under 50%-less partial feedback.

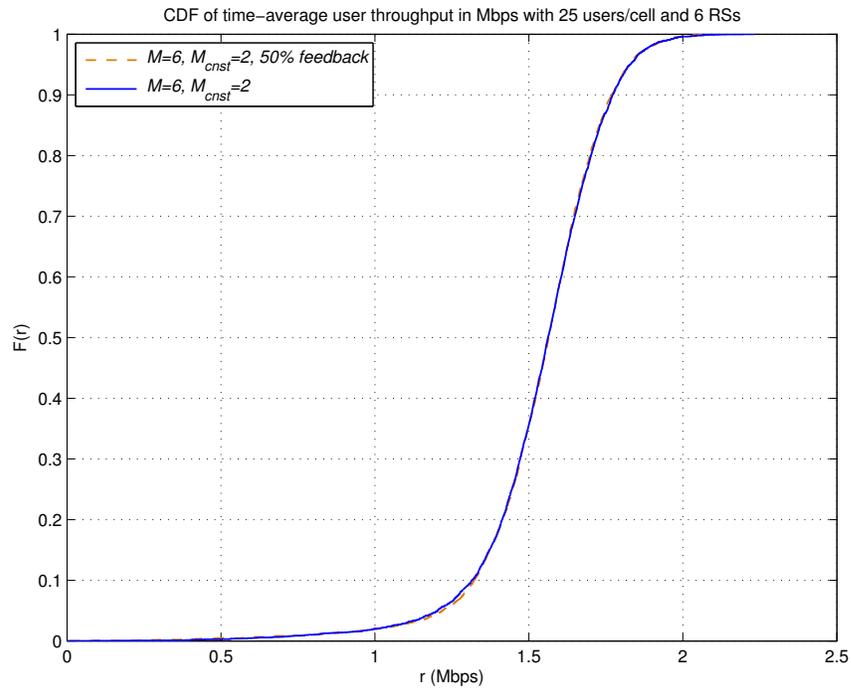


Figure 3.13: CDF of the IEEE 802.16m fairness index with 15 users/cell and 6 RSs showing no performance degradation under 50%-less partial feedback.

Based on the frame structure, the minimum delay a relayed packet encounters is $T_F + 2\tau$ where τ is the OFDM transceiver transmission time. Although the current simulation platform is quite advanced, individual packet delays are intractable. However, as the results show, the algorithm is designed to maximize throughput while stabilizing all queues and avoiding build ups. Hence, once the algorithm converges to the proper routes, it is expected to minimize the queuing delay as a consequence.

3.8 Conclusions

Efficient RRM schemes are required to harness the opportunities in the future relay-enhanced OFDMA-based networks in which user fairness is crucial. This chapter provides a novel fairness-aware joint routing and scheduling algorithm for such networks in cellular environments. The proposed algorithm exploits the opportunities in the frequency, spatial, and traffic diversities irrespective of the geographical relay deployment. As such, its performance is superior to that of a proportional fair relaying scheme in terms of ubiquitous coverage, cell-edge throughput, short- and long-term user fairness, as well as load balancing. Simulation results prove the learning ability and the efficacy of the dynamic routing strategy which converges to better routes, even under the challenging uniform relay deployment considered. The dynamic constrained routing is shown to be a more practical mode of operation in such scenarios due to the savings in feedback overhead. The inherent load-balancing feature works independently from the traffic load at adjacent cells and results as well in spatial spreading of the co-channel interference across the network. We have shown that our centralized scheme's complexity and feedback overhead are substantially less than that of traditional centralized schemes.

Chapter 4

Fair Resource Allocation Towards Ubiquitous Coverage in OFDMA Cellular Relay Networks with Asymmetric Traffic

4.1 Introduction

In Chapter 3, we have formulated a throughput-optimal policy for an OFDMA-based cellular relay network with symmetric traffic at the BS while the traditional quasi-FDR protocol of the throughput-optimal scheduling literature is retained. The policy performs joint in-cell routing and scheduling using only two-hop relaying and prevents resource waste, in contrast to prior art, through efficient bit-loading constraints or iterative optimization in the low-complexity algorithm. We have also demonstrated the system's performance in both the open routing mode and the practical constrained routing mode as compared to a relay-enhanced proportional fair scheduler (PFS). However, despite the significant performance returns of the algorithm discussed in Chapter 3, it suffers from a performance limiting bottleneck as the traffic load increases. In addition, relays in that scheme operate in the traditional quasi full-duplex mode commonly adopted in the literature on throughput-optimal scheduling. That

is, RSs are assumed to be capable of transmitting and receiving different data concurrently on orthogonal OFDM subchannels. This quasi-FDR often raises a practical concern due to the limitations in hardware technology.

Therefore, in this chapter, we present a novel throughput-optimal formulation in accordance with the emerging OFDMA-based cellular relay networks *employing half-duplex relaying*. Importantly, this chapter studies the system's performance under both *symmetric* and *asymmetric* traffic, describes the queue dynamics, and addresses the practical implementation with ARQ and constrained routing along with the potential overhead cost of queue-awareness. We first show that the network capacity, within which the policy is throughput-optimal, has been significantly increased compared to the quasi-FDR scheme, at a slight complexity increase. Hence, throughput fairness and ubiquity have been improved *at only higher traffic loads*, besides the improvement in both latency and queue-awareness.

Second, we show that without empirical priority weights, our efficient implementation of throughput-optimal scheduling achieves a ubiquitous and fair service within each class of users (with symmetric traffic) and across classes of asymmetric traffic in a relative sense, on different time scales. As such, the research contributions in this chapter can be highlighted as follows:

- A novel throughput-optimal formulation of the RRA problem in next-generation networks is developed for half-duplex relaying which is considered realistic for practical implementations [8].
- Low complexity iterative algorithms to solve the formulated optimization problem are devised where the downlink RRA over two time slots is separated using the *backlog coupling information* and the connection to the canonical end-to-end achievable capacity is introduced. Dynamic joint routing and scheduling is thus employed in contrast to most works, e.g., [43] and [106].
- We show that the network capacity for which the policy is throughput-optimal

has been significantly increased and substantial latency improvement has been achieved as compared to the quasi-FDR scheme in Chapter 3 and the prior art therein. We also explain how traffic diversity and queue-awareness are better exploited and how practical ARQ protocols can be taken into account.

- Our implementation of throughput-optimal scheduling achieves ubiquitous and fair service within each class of users (with symmetric traffic) and across classes of asymmetric traffic, on the long-term and time-average scales.
- Load balancing across relay stations (RSs) is achieved jointly with the RRA, as in Chapter 3, [86], and [9], yet only among the active RSs; no separate optimization is needed to rearrange the ‘optimal’ solution, in contrast to [44].

The journal paper [90], the patent fillings [107], and the conference paper [108] are the outcomes of the work presented in this chapter. Open research directions and possible extensions of this work are provided in Chapter 7.

4.2 System Model and Assumptions

We consider a network-level distributed/cell-level centralized RRA scheme [11], using two-hop half-duplex decode-and-forward relaying in the DL transmission of a multi-cellular network. The BS in each cell communicates with its K UTs, possibly divided into different traffic classes, directly and/or through the assistance of M fixed RSs which do not exchange traffic with each other. Based on the routing strategy, any UT may communicate simultaneously with multiple (parallel) nodes, and therefore the BS and each of the M RSs has K separate user buffers. Figure 4.1 shows a snapshot of these buffers at different cell nodes where queue lengths are represented by either blue or red bars indicating, for instance, two different inelastic traffic classes, i.e., \mathcal{K}_1 and \mathcal{K}_2 such that $\mathcal{K}_1 \cup \mathcal{K}_2 = \mathcal{K}$. This is a typical cellular setup where the traffic of a

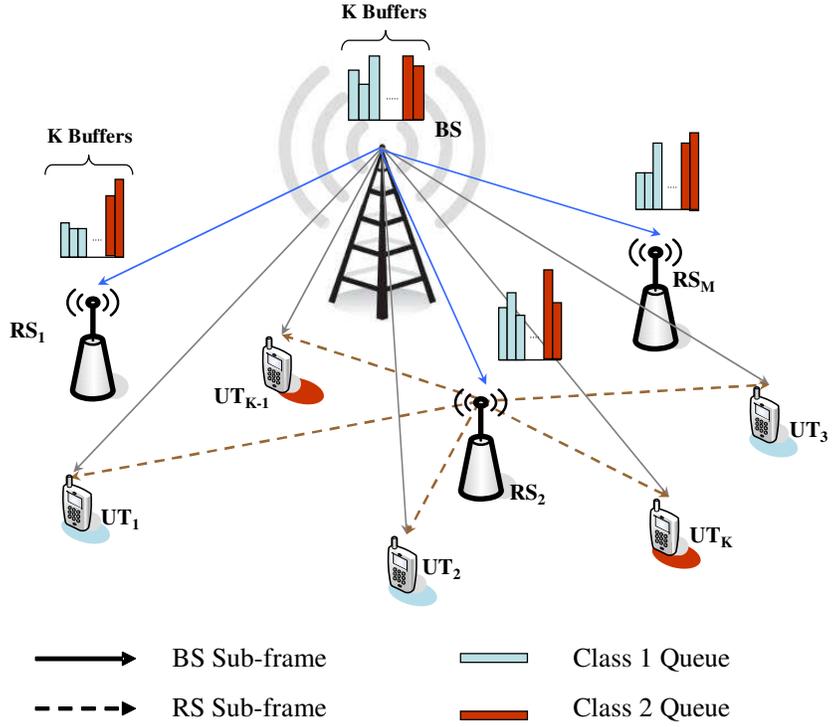
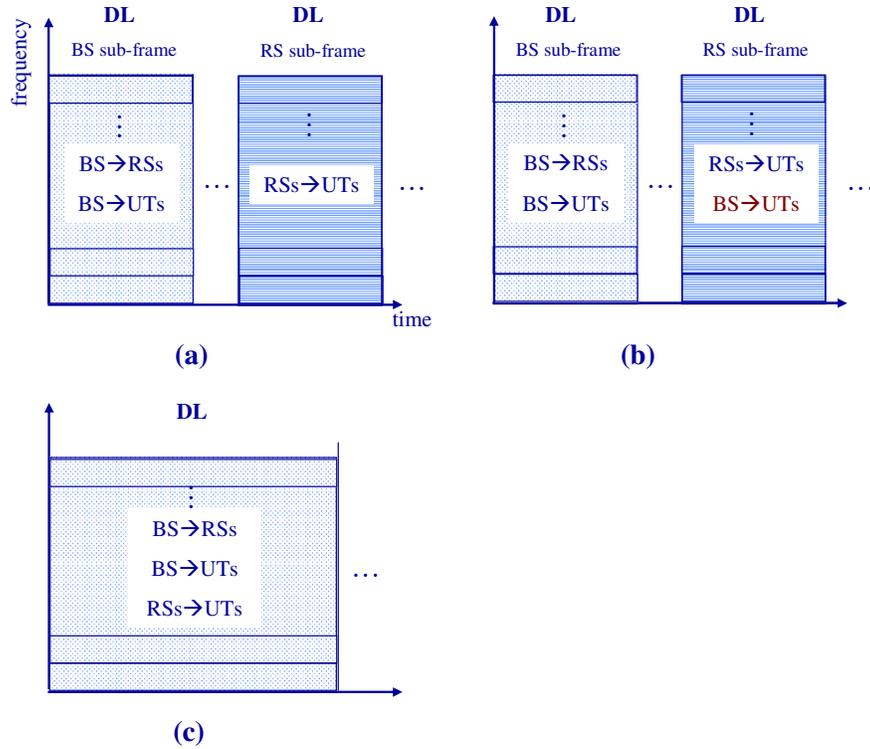


Figure 4.1: A representative cell in the multicellular network with asymmetric traffic flows and queue dynamics. The blue and red shades distinguish UTs pertaining to different classes along with their respective queues.

set of users pertaining to a certain class is generated as i.i.d following some distribution. The figure also depicts the generic operation of the joint routing and scheduling in two consecutive DL sub-frames; the BS sub-frame followed by the RS sub-frame. Aggressive resource reuse is adopted so that the same spectrum is available in each cell¹. The bandwidth is divided into N subchannels. Each subchannel is a set of adjacent OFDM data subcarriers across which the channel fading is flat. The DL frame structure of our proposed Variant-A scheme is shown in Fig. 4.2-(a) and for the sake of illustration and completeness, Fig. 4.2-(b) shows another possible protocol that defines Variant-B scheme. In any case, the coherence time of the multipath fading channel is assumed to be greater than the DL frame duration.

In the BS sub-frame (common to both Variant-A and Variant-B), only the BS transmits to the selected UTs and RSs. In the proposed Variant-A, only RSs transmit

¹Without loss of generality, this cell could resemble an LTE-Advanced ‘cell’ served by one of the three directional beams of an eNB.



5

Figure 4.2: Generic frame structures for; (a) the proposed Variant-A; (b) the investigated Variant-B; and (c) the quasi-FDR in Chapter 3.

to the selected UTs during the RS sub-frame. Whereas in Variant-B, while the RSs transmit, the BS directly transmits to some UTs who could be different from those of the first sub-frame. The sub-frame times may not necessarily be of equal length and the RRA formulation takes that into account; the sub-frame division however could be another optimization dimension that is outside the scope of this chapter. Note that according to the 802.16m frame structure in the TDD mode, for instance, the BS sub-frame is termed ‘DL Relay Zone’ and is followed first by an UL frame then by the ‘DL Access Zone’ which resembles the RS sub-frame [109]². Adaptive modulation is employed as in Chapter 3.

²Although 3GPP’s Release 10 for LTE-A has not been finalized yet, a similar scenario has been discussed in a number of recent technical reports, e.g., R1-091412 and R1-083191.

4.3 Mathematical Formulation of the RRA

In order to achieve a ubiquitous and reliable service in such systems, the RRA scheme has to dynamically route, and allocate appropriate resources to, each admitted user's traffic flow regardless of the UT's location, instantaneous traffic bursts, and short- and long-term channel conditions. In other words, the scheme has to achieve throughput fairness within each class of symmetric traffic flows, and more importantly, achieve relative fairness across these asymmetric classes such that light traffic flows are not deprived resources due to the heavy traffic flows. Due to its importance, we stress again the statements in Section 2.8.1 on our notions of fairness and ubiquity of throughput-optimal scheduling and that they are quite uncommon in the literature due to the absence of a cellular system model and/or the lack of an efficient implementation that reveals such behavior. It is also worth emphasizing that such a fairness notion does not contradict our intuition of the throughput-fairness trade off commonly observed in the literature that considers systems with continuous backlogs or full buffers. Therein, the user with the highest achievable rates would always achieve the maximum resource utilization and thus the maximum throughput, if assigned the whole resources on the expense of fairness.

Without knowledge of the channel and traffic statistics, the throughput-optimal *dynamic control policy* maximizes of the sum of the drift metric with proper constraints on frame-by-frame basis achieves our throughput and fairness objectives and exploits the degrees of freedom in multiuser, spatial, and traffic diversities as in the quasi-FDR scheme in Chapter 3. However, unlike mesh networks with multicommodity flows, all user traffic flows have to originate from only one node which is the BS in the cellular model. This indeed means that at high traffic loads, the first-hop links will create a bottleneck in the quasi-FDR scheme as the resources per DL frame have to be shared with the second-hop links which forward the previously stored data at

the RSs to the UTs, see Fig. 4.2-(c). Therefore, to improve the system's capacity, resource utilization, and to reduce the minimum delay of relayed packets, the policy has to grant the BS sole access to the resources during the first portion of the DL frame, i.e., the BS sub-frame in both Fig. 4.2-(a) and (b).

The policy uses the queue lengths at the RSs to form the differential backlog information that decides which user traffic to be routed from the BS (node 0) through RS_m , while the achievable rate $R_{0,m,n}$ on that BS- RS_m feeder link determines how many data units will be forwarded to the corresponding user's buffer at RS_m if sub-channel n is acquired. We will elaborate in Section 4.4 on the dynamics and the learning ability of this routing strategy to avoid RSs with poor access links to the destined UTs as applicable to cellular systems with only two-hop relaying. What is important to notice here that such joint policies in general deliver optimal throughput while operating as a slot-by-slot dynamic control *with no coupling* between the optimizations across time slots *except for the queue lengths*. Such an approach to routing or relay selection is also uncommon in the literature of cellular relay networks where the optimization of the frame usually involves a complex maximization of the end-to-end achievable capacity over two consecutive slots (sub-frames) as in (4.1), and thus cannot be separated into two optimizations, one per sub-frame.

$$C_{e2e}^m(k) = \frac{1}{T_1 + T_2} \min \left\{ \sum_{n_1 \in \mathcal{N}_{0 \rightarrow m, k}} R_{0,m,n_1} T_1, \sum_{n_2 \in \mathcal{N}_{m \rightarrow k}} R_{m,k,n_2} T_2 \right\}. \quad (4.1)$$

In the above, $\mathcal{N}_{m \rightarrow k}$ is the set of subchannels assigned for the access of relayed UT_k at RS_m during the RS sub-frame of duration T_2 while $\mathcal{N}_{0 \rightarrow m, k}$ is the set assigned to the feeder link during the BS sub-frame of duration T_1 .

Utilizing the flexibility in the fluid flow offered by the queue length coupling across sub-frames, two separate optimization procedures are performed (for the two sub-frames) before the BS starts transmitting in the first sub-frame. The BS sends the

allocation results for the second sub-frame to the associated RSs on separate control channels. During the uplink portion of the frame, RSs may feedback the actual status of only the affected queues as they might have changed due to the actual DL transmission. In fact, we observe that the queue length coupling information in such case is inclusive to, and more practical than, the canonical form of the end-to-end achievable capacity in (4.1) which does not consider the buffer states at the BS and accounts only for the new user data in that frame. In contrast, the queue length at the RS resulting from the first optimization q_k^m can be employed in our scheme to allocate resources to the user access link such that $\sum_{n_2 \in \mathcal{N}_{m \rightarrow k}} R_{m,k,n_2} T_2 \leq q_k^m$ and meanwhile accounts for the older data units residing at the RS that need to be rescheduled due to a practical ARQ or HARQ protocol.

4.3.1 The Joint Routing and Scheduling for the BS Sub-frame

The joint routing and scheduling optimization at the BS for the BS sub-frame can be formulated, for both variants A and B, as a binary integer linear programming (BILP) problem. As we noted earlier, the drift (demand) metric of any BS-RS feeder link on subchannel n incorporates the maximum difference between the queues at the BS and those at the RS. Also, the queue length at a UT is always zero in the DL model resembling a traffic flow sink. Therefore, the sum-demand maximization problem is formulated as

$$\begin{aligned} \max_{\boldsymbol{\rho}^{(1)}, \boldsymbol{\gamma}^{(1)}} & \sum_{n=1}^N \sum_{k=1}^K \rho_{0,k,n}^{(1)} R_{0,k,n} Q_k^0 \\ & + \sum_{n=1}^N \sum_{m=1}^M \gamma_{0,m,n}^{(1)} R_{0,m,n} \max_{k \in \mathcal{K}^m} \{(Q_k^0 - Q_k^m)^+\}, \end{aligned} \quad (4.2)$$

subject to the constraints

$$\rho_{0,k,n}^{(1)} \in \{0, 1\} \forall_{k,n}, \quad (4.3)$$

$$\gamma_{0,m,n}^{(1)} \in \{0, 1\} \forall_{m,n},$$

$$\sum_{k=1}^K \rho_{0,k,n}^{(1)} + \sum_{m=1}^M \gamma_{0,m,n}^{(1)} \leq 1 \quad \forall n, \quad (4.4)$$

$$\sum_{n=1}^N \rho_{0,k,n}^{(1)} R_{0,k,n} T_1 + \sum_{n=1}^N \sum_{m=1}^M \gamma_{0,m,n}^{(1)} R_{0,m,n} T_1 \omega_k^m \leq Q_k^0 \quad \forall k, \quad (4.5)$$

where

$$\omega_k^m = \begin{cases} 1, & k \in \arg \max_{j \in \mathcal{K}^m} \{(Q_j^0 - Q_j^m)^+\} \\ 0 & \text{otherwise.} \end{cases} \quad (4.6)$$

In (4.2)-(4.5), the first term of the objective function represents the potential users' access links directly from the BS whereas the second term represents the potential feeder links. Thus, $\rho_{0,k,n}^{(1)}$ denotes the k^{th} user's binary assignment variable to the BS on the n^{th} subchannel, during the BS sub-frame, while $\gamma_{0,m,n}^{(1)}$ is the m^{th} relay binary assignment variable to the BS node on the n^{th} subchannel ($m = 1, 2, \dots, M$ indexing the RSs). The queue length of user k at node m , expressed in bits, bytes, or packets of fixed length is denoted by Q_k^m . This queue length could change based on the allocation decisions of the BS sub-frame as in (4.7) below, and will thus be denoted by the intermediate coupling length q_k^m in the RS sub-frame's formulation. The set of all user flows that can be routed through RS_m is denoted by \mathcal{K}^m which is equal to \mathcal{K} in the open routing mode and contains only a subset of \mathcal{K} in the constrained routing mode in which, for instance, the UT provides feedback for only a preset number of the closest RSs (denoted by M_{cnst}).³ The binary indicator ω_k^m is unity if user k achieves the maximum differential backlog between the BS and RS_m , and it is 0 otherwise.

$$q_k^m = Q_k^m + \sum_{n=1}^N \gamma_{0,m,n}^{(1)} R_{0,m,n} T_1 \omega_k^m, \quad \forall m \neq 0, \quad k \in \mathcal{K}^m. \quad (4.7)$$

The constraints in (4.3) set the optimization variables to binary values while the constraints in (4.4) ensure that at most one link is active per subchannel during the BS

³Other selection criteria of relay sets could be based on pathloss rather than distance only [37].

sub-frame. Unlike the majority of works in the literature, e.g., [88], [55], and [38], the constraints in (4.5) prevent outstanding queues in this one-shot optimization from unnecessarily acquiring most, if not all, of the system resources and thus enabling throughput fairness within a class of symmetric traffic as well as across asymmetric classes. Note however that these constraints do not guarantee that a traffic flow will be allocated some or any resources at all; it is rather the role of the joint policy to maintain the stability of all the queues in the system through appropriate routing and resource allocation. As such, resource waste is also avoided and the system's capacity is therefore improved as compared to prior art. In the next section we describe the formulation of the RRA for the RS sub-frame.

4.3.2 Formulation of Variant-A for the RS Sub-frame

We recall that in the proposed Variant-A, the BS does not transmit during the RS sub-frame and only user access links are considered during that sub-frame. Here, the throughput-optimal policy operates on the coupling queue length information q_k^m which is updated by the allocation decisions of the BS sub-frame before the actual DL transmission. It is important to note that by incrementing the queues at the RSs as in (4.7) the feeder link traffic is accounted for when allocating resources to the RSs for the second sub-frame transmission. It is infeasible and noncausal, on the contrary, in the quasi-FDR scheme in Chapter 3 and the earlier literature, to account for feeder link traffic during the same DL frame due to the concurrent transmission on feeder links and RS-UT links. Therefore, the proposed implementation of throughput-optimal policies features better queue-awareness and thus in a better position towards efficient resource allocation and handling delay-sensitive traffic.

The optimization formulation of Variant-A for the RS sub-frame can be stated as

$$\max_{\boldsymbol{\rho}^{(2)}} \sum_{n=1}^N \sum_{m=1}^M \sum_{k=1}^K \rho_{m,k,n}^{(2)} R_{m,k,n} q_k^m, \quad (4.8)$$

subject to the constraints

$$\begin{aligned} \rho_{m,k,n}^{(2)} &\in \{0, 1\} \quad \forall_{m,k,n}, \\ \sum_{m=1}^M \sum_{k=1}^K \rho_{m,k,n}^{(2)} &\leq 1 \quad \forall n, \end{aligned} \quad (4.9)$$

$$T_2 \sum_{n=1}^N \rho_{m,k,n}^{(2)} R_{m,k,n} \leq q_k^m \quad \forall_{m,k}, \quad m \neq 0. \quad (4.10)$$

The binary variable $\rho_{m,k,n}^{(2)}$ assigns subchannel n to UT_k at RS_m during the RS sub-frame of duration T_2 , where $m = 1, 2, \dots, M$. Again, the constraints in (4.9) ensure that at most one link is active per subchannel during the RS sub-frame while the constraints in (4.10) caps the resources allocated to each flow to enable throughput fairness, avoid resource waste and rather achieve efficient resource utilization.

Relay fairness as appearing in the literature and discussed earlier, aims at distributing the traffic load almost evenly among RSs so that no RS will be overloaded [9]. In [86], relay fairness is assessed based on the power consumption at the RSs so that the network operates without overloading the battery of one or more RS(s). Note that if the RS's transmit power per subchannel is fixed, maintaining an almost even distribution of subchannels among RSs limits the RS's total transmit power and thus its power amplifier rating and the consumption of its battery energy, for solar/battery operated relays. This is particularly important in the context of green wireless networks where even fixed RSs rely on the solar energy. In addition, a balanced traffic load reduces the packet processing delays at the regenerative RSs and thus alleviates a practical challenge in the implementation of relay-enhanced networks. Such a feature is attained jointly with the resource allocation in this formulation by imposing the following constraint assuming uniform distribution of UTs with respect to the

geographical deployment of RSs.

$$\sum_{n=1}^N \sum_{k=1}^K \rho_{m,k,n}^{(2)} \geq \mu_A \quad \forall m \in \mathcal{M}_{act}, \quad (4.11)$$

whereas $\mu_A = \lfloor N/|\mathcal{M}_{act}| \rfloor$ is the minimum number of subchannels that should be assigned to each of the active RSs for balancing the load and $\mathcal{M}_{act} = \{m : m \neq 0, \sum_k q_k^m \neq 0\}$. Note that if $\mathcal{K}^m = \emptyset$, then $\sum_k q_k^m = 0$.

However, strict load balancing may not be desired for more practical scenarios with arbitrary distribution of UTs with respect to the RSs and with constrained routing employed since it is unlikely that RSs will handle even traffic, especially under high asymmetry between classes. Therefore, in Section 4.4, we discuss how this feature could be practically realized and integrated into our proposed iterative algorithms.

Before the following DL allocation instant, the new traffic arrivals A_k occurring within the interval between these two allocation instants are added to the user queues at the BS buffer. Whereas during the uplink frame the RSs report back their actual queue lengths to account for any variations in q_k^m due to rate deviations from the measured achievable rates $R_{m,k,n}$ or due to ARQ/HARQ requests which result in rescheduling some data units upon an erroneous reception. These dynamics can be expressed as follows assuming accurate achievable rates

$$Q_k^0 = q_k^0 + Q_{k,ARQ}^0 + A_k, \quad (4.12)$$

$$Q_k^m = q_k^m - T_2 \sum_{n=1}^N \rho_{m,k,n}^{(2)} R_{m,k,n} + Q_{k,ARQ}^m, \quad m \neq 0. \quad (4.13)$$

4.3.3 Formulation of Variant-B for the RS Sub-frame

It is clear from the literature that the transmission protocol of our proposed Variant-A scheme is not the only possible half duplex protocol. Therefore, it is interesting to observe the impact of some other transmission protocol on the performance of

our RRA formulation, especially if it provides some insight on the performance gap between the quasi-FDR protocol and the current contribution. So, in Variant-B, the BS is treated as a RS during the RS sub-frame, and thus takes a share of the resources to communicate directly with some selected UTs. As such, the formulation of Variant-B scheme during the RS sub-frame is the same as that of Variant-A but with the node index m ranging from 0 to M in the equations (4.8) through (4.10). The queue length dynamics as a result of the DL transmission and before the following allocation instant will still follow (4.13) for the RSs whereas (4.14) applies for the queues at the BS.

$$Q_k^0 = q_k^0 - T_2 \sum_{n=1}^N \rho_{0,k,n}^{(2)} R_{0,k,n} + Q_{k,ARQ}^0 + A_k. \quad (4.14)$$

Having stated our novel RRA problem formulation, achieving the aforementioned objectives, as the authors stress in [56], depends on devising efficient and practical algorithms to realize the proposed schemes as applied to the cellular system and subject to the associated set of constraints. Since the computational complexity of the above BILP formulation is $\mathcal{O}\left((M_{cnst}K)^N\right)$, $M_{cnst} \leq M$, and given the expected numbers of subchannels, UTs, and RSs in a practical cellular network, it is inevitable to devise suboptimal low-complexity iterative algorithms to circumvent the prohibitive complexity levels. Therefore, we propose the following iterative algorithms to solve the formulated optimization and achieve its throughput and fairness objectives with tolerable polynomial complexity.

4.4 Realization of the Schemes Through Low-Complexity Iterative Algorithms

The BS sub-frame allocation procedure is the same for both Variant-A and Variant-B. The BS has full access to all of the N subchannels yet the transmission occupies only a portion of the DL frame duration (see Fig. 4.2). The demand metric of any BS-UT link on subchannel n is given as

$$D_{n,0 \rightarrow k} = R_{0,k,n} Q_k^0, \quad (4.15)$$

and the demand metric of any BS-RS link on subchannel n is expressed as

$$D_{n,0 \rightarrow m} = R_{0,m,n} \max_{k \in \mathcal{K}^m} \{(Q_k^0 - Q_k^m)^+\}. \quad (4.16)$$

We denote the destination of the ‘best’ BS link, i.e., with the maximum demand, out of $K + M$ potential links on subchannel n , as \hat{j}_n . The algorithm then finds the highest demand across all the unassigned subchannels and the associated BS link denoted as \hat{j} is then selected. The algorithm runs another iteration after eliminating the assigned subchannel and updating the associated queue(s). The iterative process stops when subchannels are exhausted or the queues at the BS are evacuated. Using this greedy iterative assignment approach the sum-demand is maximized in compliance with the constraints (4.3) through (4.4) while the efficiency and fairness-enabling constraints (4.5) are satisfied by updating the affected queue(s) (at the BS and the RSs if applicable), according to the assigned rates. Therefore, the BS queues with high traffic load are given their natural priority and allocated the subchannels with the highest achievable rates until they come to around the same back pressure of the low traffic queues, then the joint policy uses the remaining subchannels to stabilize all the BS queues. Note that in the literature on throughput-optimal policies, a CAC

mechanism is usually assumed at a higher level to either grant or deny any of these traffic flows service based on the system's capacity [56]. The CAC mechanism has been integrated with a weight-based scheduler in [110] to provide a differentiated service to the heterogenous queues but in a conventional cellular network. In that context, a design of fair weights has been also presented in [111] based on utility proportional fairness rather than rate proportional fairness.

In the following we present the pseudo-codes of the RRA algorithms for the BS sub-frame and the RS sub-frame based on Variant-A. In these codes, \mathcal{U} , \mathcal{N} , \mathcal{K} , and \mathcal{M} denote the sets of unassigned subchannels, all available subchannels, UTs, and RSs, respectively. Recall that in Variant-A (as explained partly in Fig. 4.2), the BS does not transmit at all and only RSs share the resources to transmit to the selected UTs during the RS sub-frame. Similar to the BS-UT links in the previous algorithm, the algorithm here finds in each iteration the best link from any relay RS_m , out of the $|\mathcal{K}^m|$ links to UTs, on the unassigned subchannel n ; such maximum is denoted by $D_{n,m}$.⁴ Since only one link will be active per subchannel, the algorithm needs to compare $D_{n,m}$ across all RSs for each subchannel. If the load balancing constraints are not imposed, then on each iteration the algorithm assigns subchannel \hat{n} to the best link from $\text{RS}_{\hat{m}}$; i.e., line (10) in the second pseudo-code is replaced by

$$(\hat{n}, \hat{m}) = \arg \max_{n,m} D_{n,m}. \quad (4.17)$$

However, if the load balancing constraints are imposed, the algorithm solves an optimal one-to-one assignment problem per iteration to maximize the total demand by applying the Hungarian algorithm [101] to the tall $|\mathcal{U}| \times M$ demand matrix $[D_{n,m}]$.

⁴Since there is no interdependency between the links at different (n, m) pairs, maximizing over k for each pair (n, m) does not affect the combinatorial problem, i.e., does not change the optimal solution.

After each iteration, user queues are updated based on the assigned rates. The iteration process continues until all the traffic in the RS queues is scheduled or the subchannels are exhausted. Note that our implementation of the Hungarian algorithm excludes in any iteration the columns (RSs) with all zero entries; this occurs when the RS has no further traffic to be scheduled or it did not receive any traffic at all, i.e., $\mathcal{K}^m = \emptyset$. As such, when the combined load, due to high and low traffic flows, is almost uniform across all the RSs, the one-to-one assignment jointly achieves strict load balancing and equal power consumption among RSs. Whereas, when some RSs are inactive and some are handling higher traffic loads than the others, that same algorithm will maintain, from iteration to another, the even distribution of subchannels among only the active RSs including the lightly loaded ones which eventually turn inactive and then the balancing continues in the remaining iterations among the heavily loaded RSs, and the process continues. Such flexibility in the proposed algorithm, due to the iterative Hungarian, makes it more suitable for practical scenarios as the load is autonomously balanced in a relative sense without invoking an additional optimization.

Modifying the RS sub-frame algorithm to employ the half-duplex protocol of Variant-B is done by simply running the node index m from 0 to M and thus the dimension of the demand matrix in any iteration becomes $|\mathcal{U}|$ -by- $(M + 1)$.

4.4.1 Dynamic Routing in the Two-hop Cellular Relay Network

As we discussed earlier in Chapter 3, routing in the context of mesh networks employing throughput-optimal policies is performed dynamically using the maximum differential backlog from node a to node b, $\max_k \{Q_k^a - Q_k^b\}$, and the route may comprise an indefinite number of hops. This is undesirable and expensive, especially in cellular networks operating in licensed bands. It is not also realistic to assume knowledge of the CSI between any arbitrary pair of RSs across all subchannels; and

Pseudo-code for the BS sub-frame for both variants

1. Initialization: $\mathcal{U} = \mathcal{N}$, update $\mathbf{Q}^0 = [Q_1^0 \dots Q_K^0]$ by new arrivals \mathbf{A} , update affected queues in \mathbf{Q}^m by feedback and ARQ rescheduling $\mathbf{Q}_{\text{ARQ}}^m$.
 2. **while** $|\mathcal{U}| \neq 0$ and $\mathbf{Q}^0 \neq \mathbf{0}$ **do**
 3. **for** each $n \in \mathcal{U}$
 4. **for** $m = 1$ to M
 5. $D_{n,0 \rightarrow m} = R_{0,m,n} \max_{k \in \mathcal{K}^m} \{(Q_k^0 - Q_k^m)^+\}$
 6. $\kappa^m = \arg \max_{k \in \mathcal{K}^m} \{Q_k^0 - Q_k^m\}$
 7. **end for**
 8. **for** $k = 1$ to K
 9. $D_{n,0 \rightarrow k} = R_{0,k,n} Q_k^0$
 10. **end for**
 11. $D_{n,0} = \max_j \{D_{n,0 \rightarrow j}\}, \quad j \in \mathcal{K} \cup \mathcal{M}$
 12. $\hat{j}_n = \arg \max_j \{D_{n,0 \rightarrow j}\}$
 13. **end for**
 14. $\hat{n} = \arg \max_n \{D_{n,0}\}, \mathcal{U} = \mathcal{U} - \{\hat{n}\}, \hat{j} = \hat{j}_{\hat{n}}$
 15. **if** $\hat{j} \in \mathcal{M}$ **then**
 16. $\hat{k} = \kappa^{\hat{j}}, \quad b = \min\{Q_{\hat{k}}^0, \lfloor R_{0,\hat{j},\hat{n}} T_1 \rfloor\}$
 17. $Q_{\hat{k}}^0 = Q_{\hat{k}}^0 - b, \quad Q_{\hat{k}}^{\hat{j}} = Q_{\hat{k}}^{\hat{j}} + b$
 18. **else**
 19. $\hat{k} = \hat{j}, \quad Q_{\hat{k}}^0 = (Q_{\hat{k}}^0 - \lfloor R_{0,\hat{k},\hat{n}} T_1 \rfloor)^+$
 20. **end if**
 21. **end while**
-
-

Pseudo-code for RS sub-frame for Variant-A

1. Initialization: $\mathcal{U} = \mathcal{N}$, $\mathbf{q}^m = \mathbf{Q}^m \forall m$.
 2. **while** $|\mathcal{U}| \neq 0$ and $\sum \mathbf{q}^m \neq \mathbf{0}$ **do**
 3. **for** each $n \in \mathcal{U}$
 4. **for** $m = 1$ to M
 5. $D_{n,m} = \max_k \{R_{m,k,n} q_k^m\}$
 6. $\kappa_{n,m} = \arg \max_k \{R_{m,k,n} q_k^m\}$
 7. **end for**
 8. **end for**
 9. % $\mathbf{D} = [D_{n,m}]$ is the demand matrix.
 10. $(\hat{\mathbf{n}}, \hat{\mathbf{m}}) \leftarrow \mathbf{Hungarian}(\mathbf{D})$ % Vectors of indices
 11. $\mathcal{U} = \mathcal{U} - \{\hat{\mathbf{n}}\}$, $N_{assigned} = |\hat{\mathbf{n}}| = |\hat{\mathbf{m}}|$
 12. % $N_{assigned} \leq \min\{M, |\mathcal{U}|\}$
 13. **for** $i = 1$ to $N_{assigned}$
 14. $\hat{n} = \hat{\mathbf{n}}(i)$, $\hat{m} = \hat{\mathbf{m}}(i)$, $\hat{k} = \kappa_{\hat{n}, \hat{m}}$
 15. $q_{\hat{k}}^{\hat{m}} = (q_{\hat{k}}^{\hat{m}} - \lfloor R_{\hat{m}, \hat{k}, \hat{n}} T_2 \rfloor)^+$
 16. **end for**
 17. **end while**
-
-

it is also unlikely with uniform relay deployment that all RSs have good links to the UT. Therefore, we restrict the dynamic routing to the commonly adopted setup, i.e., two-hops at most, and thus RSs are not allowed to exchange traffic. Hence, the differential backlog terms take the form $\max_{k \in \mathcal{K}^m} \{Q_k^0 - Q_k^m\}$, where $\mathcal{K}^m = \mathcal{K}$ in the hypothetical open routing mode (any UT may receive from any RS) and $\mathcal{K}^m \subseteq \mathcal{K}$ in the practical constrained routing mode (only the best RSs are considered for a UT).

Consequently, in the open routing mode, initial accumulation of the user's traffic may occur at some RS(s) with poor links to the UT as such traffic will neither be forwarded to the UT nor will it be absorbed by another RS. However, the maximum differential backlog exploits the presence of the trapped data at these RSs, indicating the quality of the second-hop links, and reduces the likelihood of forwarding the user's data on such feeder links in following iterations and allocation instants. In Chapter 3, under uniform deployment of RSs and with different M_{cnst} , the narrow performance gap between the open and constrained routing modes of the quasi-FDR algorithm demonstrates this inherent learning ability of the routing strategy to avoid routes with poor second hops in the open mode. We stress that the improvement due to constrained routing comes along with substantial savings in feedback overhead due to the eliminated links as discussed in Section 4.6. This learning ability of the joint strategy is also inherent in the proposed algorithms in this chapter. So, besides the fairness and ubiquity aspects across asymmetric traffic flows, this observation on the routing behavior of throughput-optimal policies as applied to two-hop cellular relay networks is also quite interesting since the common understanding is that imposing constraints on the routing options might reduce the capacity of the multicommodity mesh network.

4.4.2 The Computational Complexity

The computational complexity of Variant-A and Variant-B schemes discussed in this chapter is found to be polynomial in time of $\mathcal{O}(\frac{N^2(N+M)^2}{4M})$, $M \leq N$, and $\mathcal{O}(\frac{N^2(N+M+1)^2}{4(M+1)})$, $M + 1 \leq N$, respectively. These estimates come from the fact that the complexity of the Hungarian algorithm is of $\mathcal{O}(|\mathcal{U}|^3)$. These are the complexity levels incurred in the second sub-frame. However, the proposed scheme incurs a slight increase in complexity of $\mathcal{O}\left(\frac{N^2}{2}(K + M)\right)$ due to the first sub-frame allocation as compared to the iterative algorithm of the quasi-FDR reference scheme in Chapter 3.

4.5 Numerical Results

The network and channel parameters used in this study are given in Table 3.1. Most of the parameters are taken from the 3GPP LTE release 9 (Case 3) [112] or the WiMax Forum [105] while the WINNER C2 channel model [92] is used. To allow for a fair comparison, the same setup of Chapter 3 has been considered with 19 hexagonal cells enhanced with 3 or 6 relays, with equal angular spacing, in each cell. The total DL frame length is 2 msec with equal sub-frame durations ($T_1 = T_2$). The UTs in each cell are uniformly distributed over the cell area. Since throughput-optimal policies can be applied regardless of the traffic and channel distributions, independent Poisson packet arrival processes are assumed at the BS queues. The average arrival rate for a Class 1 UT is $\lambda_1 = \lambda$ and for a Class 2 UT is $\lambda_2 = 2\lambda$ where λ is 632 packets (188 bytes each) per second.

On top of the 4-dB lognormal shadowing, the BS-RS links experience *time-frequency correlated Rician* fading with a Rician factor of 10 dB. All other links are NLOS and experience 8.9 dB independent lognormal shadowing with *time-frequency correlated Rayleigh fading*. Each RS employs an omni-directional antenna to communicate with

UTs and a highly directive receive antennas with a horizontal gain pattern given in [112] to communicate with its BS. The user mobility used for the study is 20 Km/hr, however the scheme can support mobility as high as 90 Km/hr, given the frame structure and the resulting channel coherence time. In each drop, user locations and shadowing realizations are maintained constant for which the subject schemes need to compensate while the traffic and the channel vary on frame-by-frame basis as time evolves.

4.5.1 The Proposed Scheme vs. the Quasi-FDR scheme and Prior Art

We first consider the case where all UTs belong to the same class, say Class 1, and the cellular network thus handles all symmetric traffic flows ($\mathcal{K}_1 = \mathcal{K}$). Figure 4.3 shows CDF plots of the time-average user throughput across all drops with $K = 30$ UTs and $M = 3$ or 6 RSs per cell. The same amount of resources are provided for all schemes, i.e., same DL frame length, bandwidth, and total transmit powers. The figure shows that the quasi-FDR scheme, even in its open routing mode, outperforms the channel-aware only relay-enhanced proportional fair scheme (PFS), which is discussed in Chapter 3, and shows that at that loading level, a significant throughput gain is realized with the proposed Variant-A scheme indicating a bottleneck in the quasi-FDR. The figure also shows that the performance gap increases as the number of RSs increases; this can be attributed to the capability of the proposed scheme, as opposed to the quasi-FDR at that point, to exploit the potential increase in spatial diversity and thus in the system's capacity when more RSs are deployed with closer proximity to the UTs and good feeder links.

This is in line with our understanding that the bottleneck results from the BS taking only a share of the resources to transmit directly to some UTs as well as forwarding the relayed traffic to the feeder links. As will be shown in Fig. 4.5, this does not limit the performance at light to moderate loadings, whereas at higher

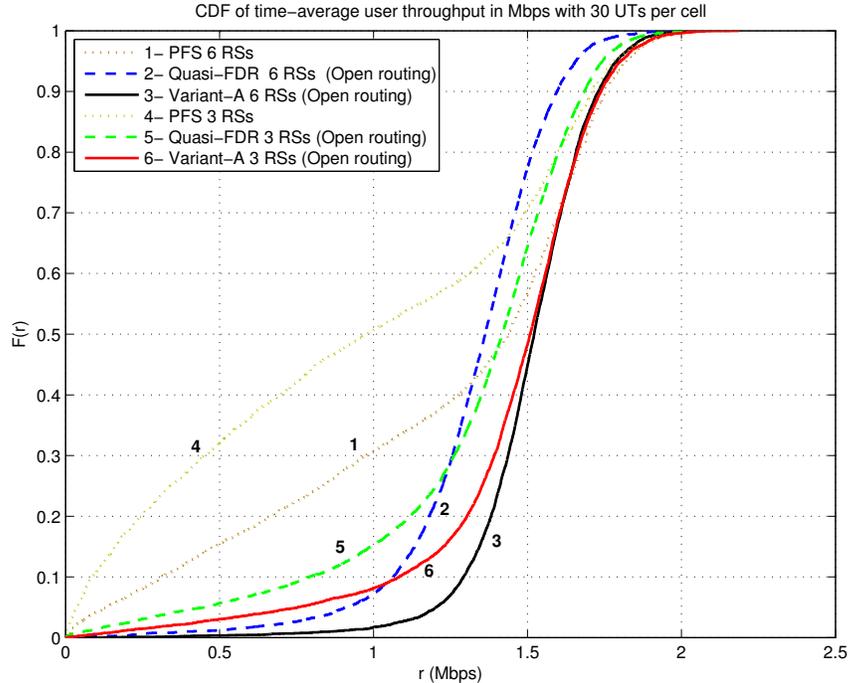


Figure 4.3: Time-average throughput comparison of Variant-A with the reference schemes at 30 UTs/cell.

loading, this share of resources becomes insufficient to serve the traffic. On the other hand, the proposed scheme in this chapter grants the BS full access to the whole bandwidth during the BS sub-frame.

Another informative way of reading these results, according to the LTE evaluation methodology [112], is comparing the cell-edge user throughput attained by the schemes at the 5th percentile. The zoom-in window on the lower tail behavior in Fig. 4.4 shows that our proposed scheme yields a superior cell-edge performance over the quasi-FDR at a much higher load ($K = 40$). The CDFs of the quasi-FDR scheme with 3 and 6 RSs show that reducing the number of RSs relieves the bottleneck at that load to some extent (by increasing the resource share of the BS) and thus improving the upper tail. However, the spatial diversity required to enhance the cell-edge throughput is lost and thus affecting the throughput fairness as shown in Fig. 4.6 and Fig. 4.7.

The time-average fairness performance of the proposed scheme is presented in Fig. 4.6 for $K = 40$ using CDF plots of the fairness metric in [87] which can be

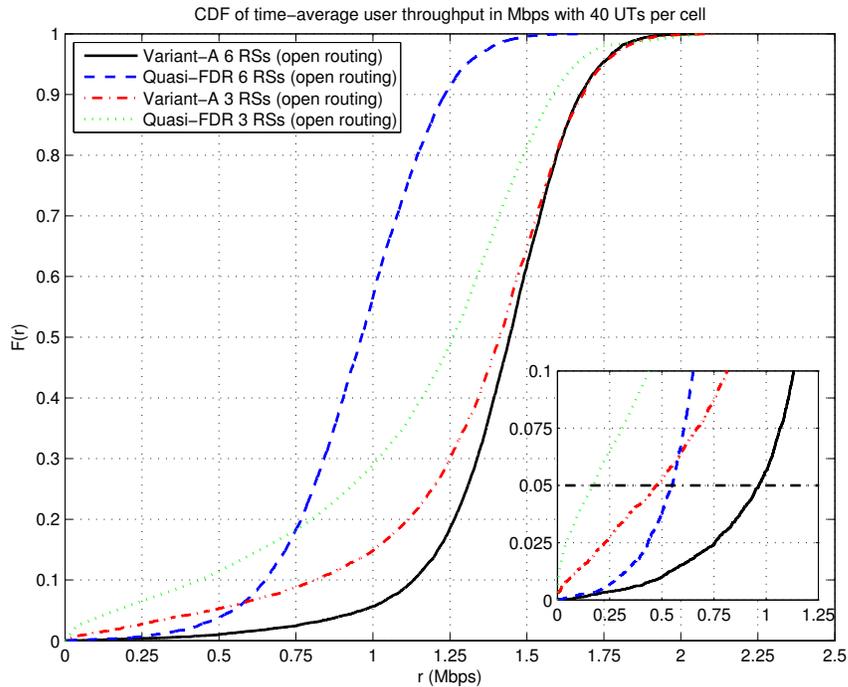


Figure 4.4: The CDF of time-averaged user throughput with 40 UTs per cell and emphasis on the lower tail behavior.

defined as in (4.18) with $\beta_i = 1 \forall i$. Using Jain's index [47] as defined in (4.19), the long-term fairness is demonstrated for $K = 40$ in Fig. 4.7 where $r_{i,w}$ is the throughput of UT_i during a time window w of 20 frames and $\beta_i = 1 \forall i$. In these fairness figures, a step function at unity in the CDF plots indicates absolute fairness. Therefore, the closer the curve is to a step function at unity the more fair the scheme is. It is observed therefore that the proposed scheme achieves the most fair performance as compared to the reference scheme, in the time-average sense and even in the long-term sense. This further underscores the superiority of the proposed scheme in highly loaded networks.

Figure 4.5 shows the average total cell throughput as a function of the number of UTs/cell. It is clear that the performance gap between the two variants of the half-duplex scheme and the quasi-FDR scheme increases significantly as the load increases and becomes insignificant at low to moderate loading levels. The impact of the bottleneck in the quasi-FDR scheme can be realized by comparing the slope of these curves at the high loading end where the proposed scheme outperforms the

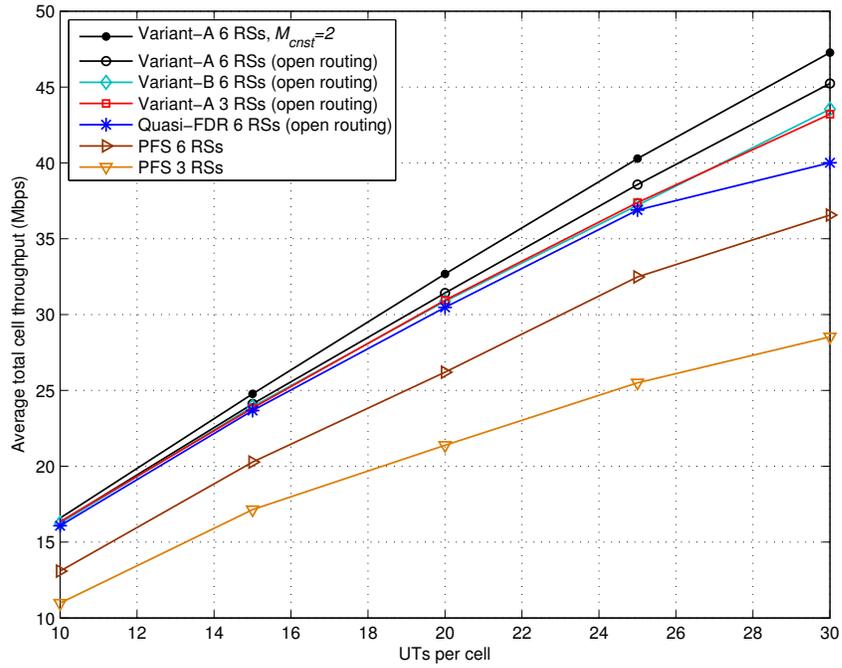


Figure 4.5: The total cell average throughput vs. the number of UTs per cell for the proposed and reference schemes.

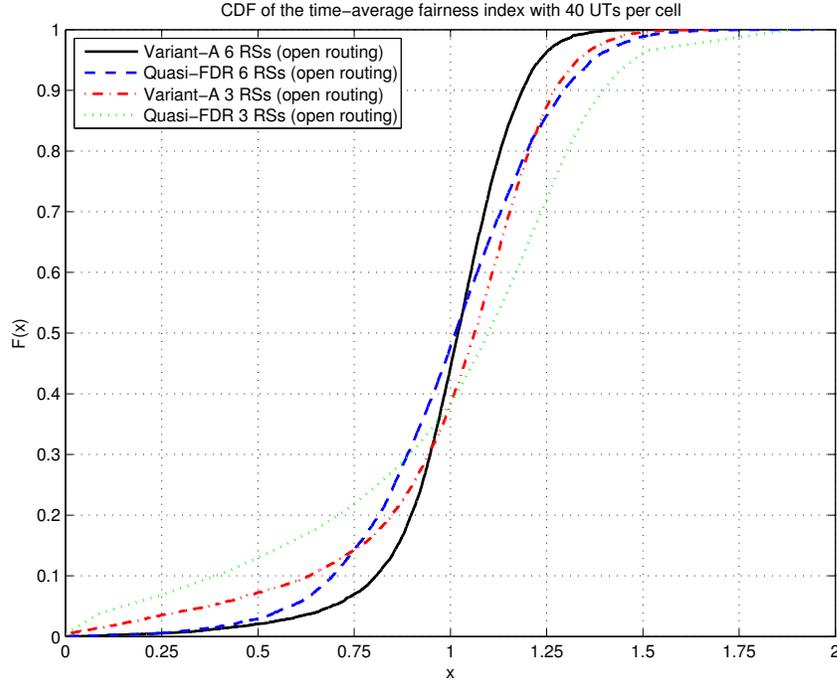


Figure 4.6: Time-average throughput fairness for the proposed scheme and the reference quasi-FDR scheme with 40 UTs per cell.

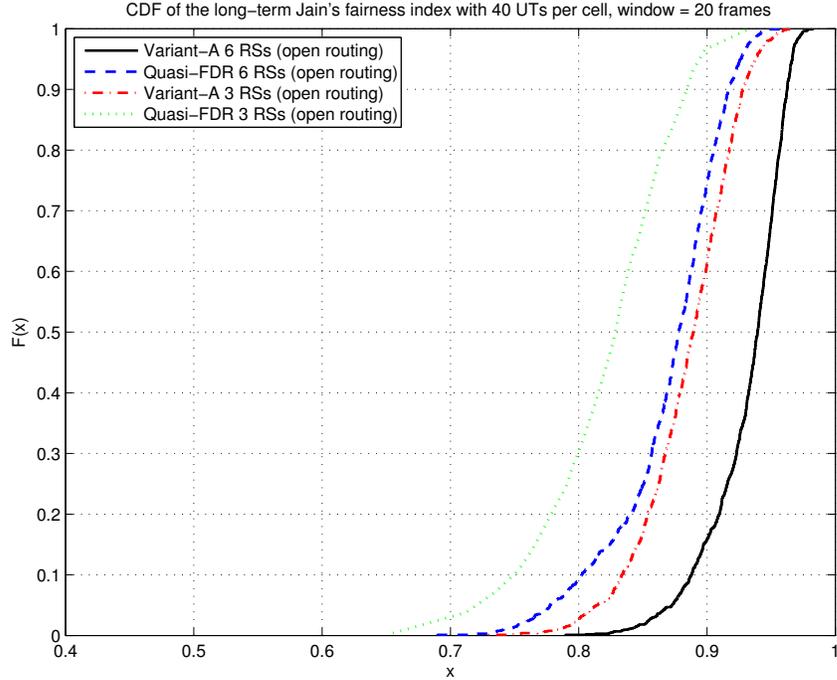


Figure 4.7: Long-term throughput fairness for the proposed scheme and the reference quasi-FDR scheme with 40 UTs per cell and a time window of 20 frames.

reference one even with fewer RSs. It is worth mentioning that in Variant-B, more resources are devoted to the BS than in Variant-A due to the allocation in the RS sub-frame. However, this results in the scheduling policy becoming less flexible, as the load increases, forwarding the relayed traffic to destined UTs. The performance of Variant-B with 6 RSs is almost the same as that of Variant-A with 3 RSs, yet it is superior to that of the quasi-FDR with 6 RSs. As discussed earlier in Chapter 3, in addition to its substantial feedback savings, constrained routing in two-hop cellular networks enables the joint routing and scheduling policy to achieve a better performance exploiting the deployment geography; this is demonstrated here by the top curve in this figure representing Variant-A with 6 RSs but using $M_{cnst} = 2$ closest RSs. As such, throughout the rest of our results, the proposed half-duplex scheme will be represented by Variant-A with $M_{cnst} = 2$. Figure 4.5 also shows that the relay-enhanced PFS is significantly inferior to all other schemes; this is due to the

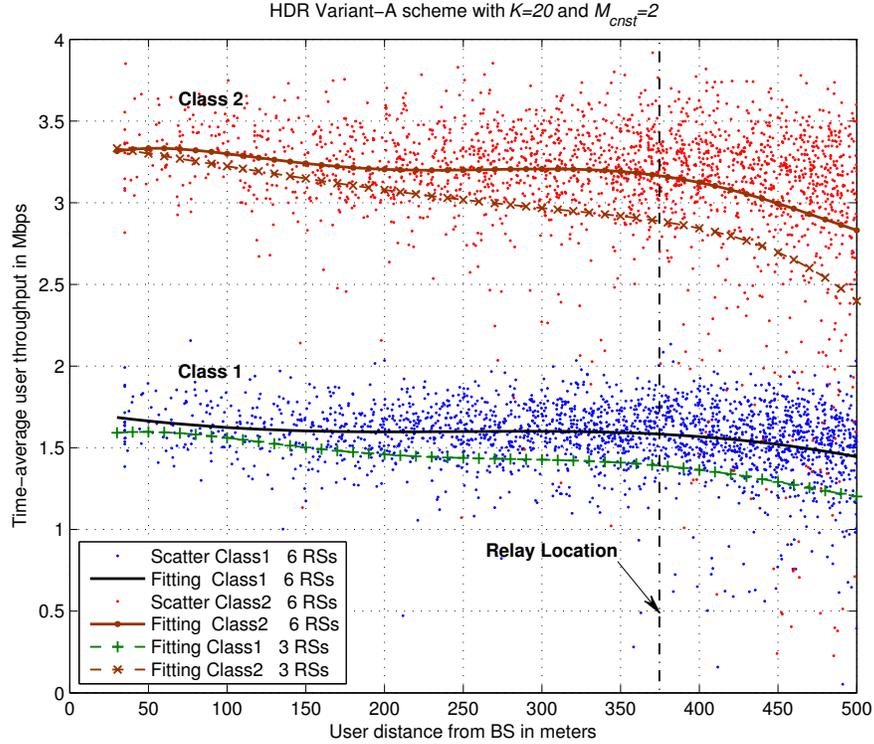


Figure 4.8: Time-average user throughput as function of user location and shadowing with 20 UTs/cell with asymmetric traffic and 3 or 6 RSs. Other scatters are not shown for figure clarity.

lack of traffic- or queue-awareness and the partitioning of resources and UTs which is commonly adopted in literature. A comparison with other non-relaying schemes can be found in [108].

4.5.2 Variant-A with symmetric and asymmetric traffic

We now consider the case where the UTs are equally divided into two groups \mathcal{K}_1 and \mathcal{K}_2 , namely Class 1 and Class 2, and the cellular network thus handles asymmetric traffic flows with $K_1 = K_2 = K/2$. The aggregate offered traffic load is represented by the aggregate mean arrival rate $\Lambda = \lambda K/2 + 2\lambda K/2$ which is the aggregate load when all UTs belong to Class 1 such that $K' = 3K/2$. The latter scenario is thus used as a reference scenario given the same resources and number of RSs.

Figure 4.8 shows a scatter plot of user time-averaged throughput as a function of user distance from the BS using Variant-A with $K = 20$, $M = 3$ or 6, and $M_{cnst} = 2$. Each point in the scatter represents the time-averaged throughput for a particular

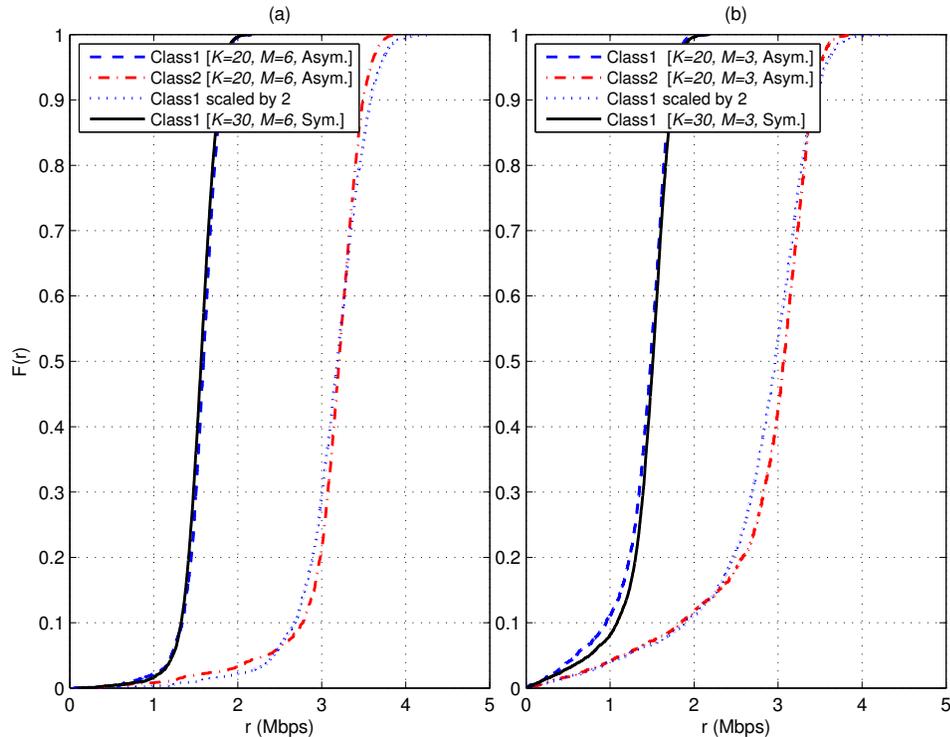


Figure 4.9: The CDF of time-averaged user throughput with symmetric and asymmetric traffic.

UT within a drop with fixed location and shadowing. The location-based conditional mean is approximated by a 5th degree polynomial curve fitting as a means of averaging out the effect of shadowing on the joint policy. The figure indicates that uniform average throughput across the cell area is achieved and thus a ubiquitous service is attained within each of these asymmetric traffic classes without imposing priorities on the RRA formulation. This is deduced from the almost flat fittings and the confined spreading of the scatter points. In line with our understanding of the impact of RSs on the capacity and cell-edge performance, the fittings with 3 RSs show less ubiquity and less cell-edge throughput as compared to the case with 6 RSs. In both cases, it can be observed that relatively more spreading of the scatter points and less cell-edge throughput are realized for Class 2 UTs as compared to Class 1 UTs.

CDF plots of the scatter points with 6 and 3 RSs are shown in Fig. 4.9-(a) and (b), respectively. The lower tail behavior attests to the latter observation on the relative

cell-edge performance between Class 1 and Class 2 UTs. To have some insight on the relatively higher spreading (or variance) of Class 2 points, a hypothetical CDF plot is generated by scaling up the time-average throughput realizations of Class 1 UTs by $\beta = 2$. We can generally define the normalizing factor for flow i as $\beta_i = \lambda_i/\lambda_1$. The concordance between the hypothetical CDF and that of Class 2 especially in terms of variance, and to some extent slope, reveals that the proposed scheme provides almost the same service to the asymmetric traffic flows but in a relative sense, i.e., the realizations of Class 2 service could be roughly approximated by a transformation of the realizations of Class 1 service using the scaling factor β .

Comparing the CDF of Class 1 UTs in the case of asymmetric load ($K = 20$) to that of Class 1 in the reference case of symmetric load ($K' = 30$), it is observed that with 6 RSs, the reference curve has an insignificant improvement, mainly at the lower tail. Note that the potential for improvement should be attributed to the increased multiuser/frequency diversity at $K' = 30$. Despite its coexistence with Class 2 traffic, Class 1 traffic receives similar service to that in the all symmetric case, given the same aggregate load and the same resources. However, with 3 RSs and thus less spatial diversity, the improvement with $K' = 30$ becomes more visible.

The corresponding time-average fairness performance of the previous cases is shown in Fig. 4.10 using (4.18) with $\beta_i = 1 \forall i$ for absolute fairness within the same class and with the normalized throughput, as in [87] and [85], for relative fairness across the asymmetric classes.

$$x_j = \frac{r_j/\beta_j}{\frac{1}{K} \sum_{i=1}^K r_i/\beta_i}. \quad (4.18)$$

Similarly, the long-term fairness is shown in Fig. 4.11 using Jain's index in (4.19) with $\beta_i = 1 \forall i$ for absolute fairness within the same class and using the normalized

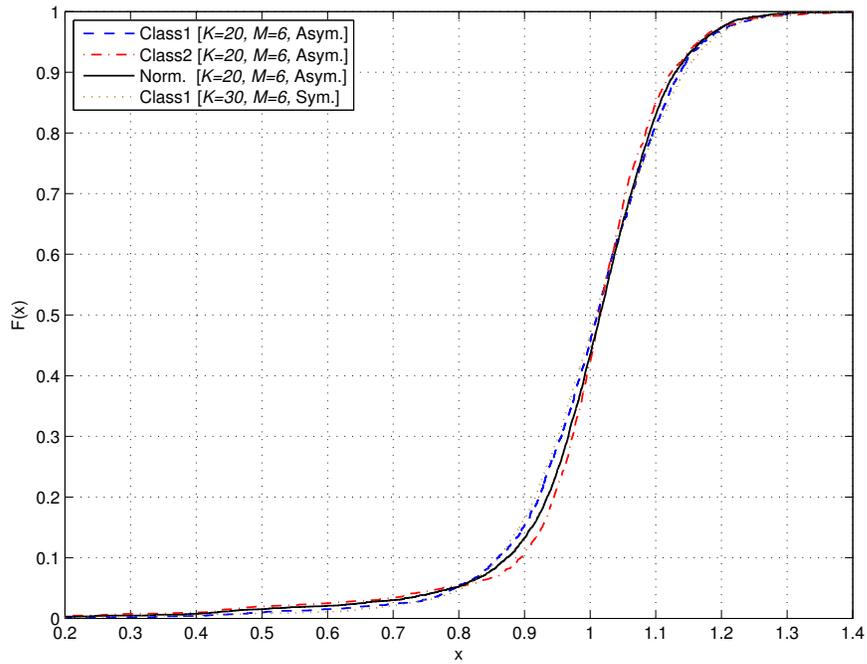


Figure 4.10: The time-average absolute and relative fairness with symmetric and asymmetric traffic.

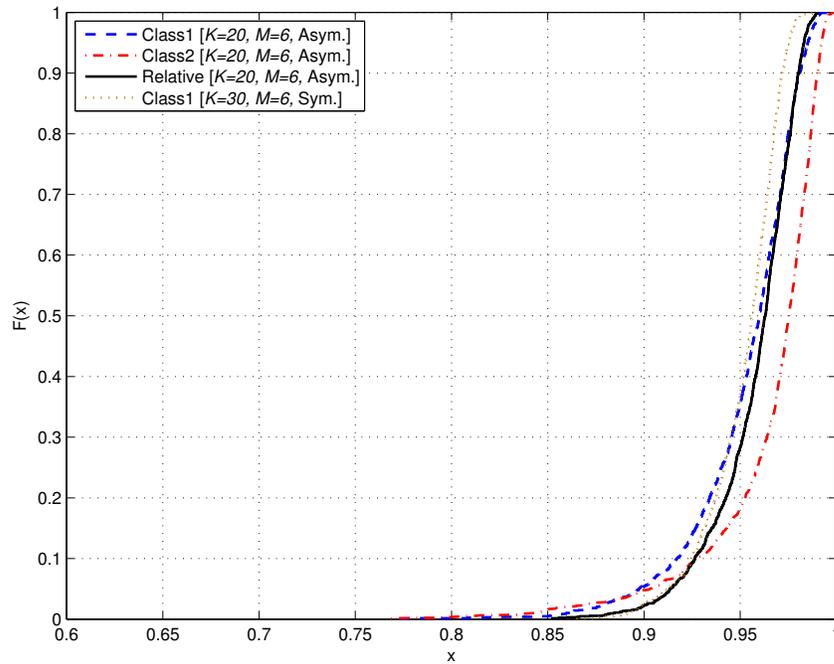


Figure 4.11: Long-term absolute and relative fairness with symmetric and asymmetric traffic.

throughput as in [84] for relative fairness.

$$x_w = \frac{\left(\sum_{i=1}^K r_{i,w}/\beta_i\right)^2}{K \sum_{i=1}^K (r_{i,w}/\beta_i)^2}. \quad (4.19)$$

In general, all class-based absolute fairness curves are quite close while the relative (normalized) fairness curve also lies in between. The slight improvement in the absolute fairness of Class 2 can be attributed to the lower sensitivity of the fairness functions at high rate values along with the slight throughput improvement shown in the CDF of Fig. 4.9 over the scaled up throughput of Class 1. This further underscores the superiority of the half-duplex scheme in highly loaded networks. Once again, the performance of the reference case with $K' = 30$ and 6 RSs matches that of Class 1 in the asymmetric case.

4.6 Implementation Issues and Feedback Overhead

In contrast to traditional cell-level centralized RRM schemes, substantial savings in CSI feedback overhead can be achieved due to the following reasons which are discussed in more detail in Chapter 3: 1- Implementing the constrained routing mode reduces the feedback overhead by a factor of $(M_{cst} + 1)/M + 1$ since no feedback is required from the UT for the eliminated RS-UT links. 2- Reporting the indices of the achievable AMC levels per link significantly saves in signalling overhead as compared to reporting a wide range of continuous SINRs. 3- Having many UTs per cell and given that a UT can be connected to more than one node, only the ‘best’ fraction (in terms of achievable rates) of the N subchannels needs to be feedback per user access link; this reduces the overhead by a factor of N_{CSI}/N .⁵

⁵Results considering only the best 50% of link subchannels in Chapter 3 show no performance degradation.

Since queue-awareness at the BS is a key element in the proposed RRA algorithms, it is important to investigate whether or not there is an associated overhead cost as compared to channel-aware-only relay-based schemes. Looking at the queue length dynamics described in (4.12) to (4.14), it can be realized that the required queue state results from updating the former state by the new traffic arrivals, the RRA decisions, and finally, the ARQ rescheduling requests whose overhead is neuter in this investigation. As such, we observe that the BS can update the queue length information about its cell nodes spontaneously (at no cost), or at a minimal cost if system design necessitates, due to the following reasons:

1. Since UTs are the flow sinks of the DL traffic, their queue lengths are set to zero without incurring an overhead cost. Whereas the BS is self-aware of its full queue dynamics including the ARQ requests from the former recipients of its transmissions.
2. In contrast to mesh networks, new traffic arrivals occur only at the BS node in the cellular network which implies that no exogenous arrivals at RSs or UTs need to be reported to the BS.
3. Since the RRA is cell-level centralized, the BS is aware of the data transmitted from the RSs to the UTs. Whereas, the relayed data withdrawn from the BS buffers is used by the BS to increment the queue images of the destined RS(s).
4. If a UT generates an ARQ to an RS, the protocol may enable the BS to exploit the broadcast channel to infer the amount of data incrementing back the RS queue and hence update the corresponding image accordingly. If the system design necessitates otherwise, or rules out ARQ while channel impairments may cause data losses, then during UL, the RS will need to report the actual change in *only the queues affected by the last DL transmission, over its potentially high-speed feeder link*⁶.

⁶Furthering the savings in overhead, some quantization of the queue length process, expressed

It is worth noting that the proposed algorithms exploit the finer resource granularity of the half-duplex frame structure despite the slight increase in complexity due to the BS sub-frame. However, at low to moderate loading levels, the quasi-FDR scheme achieves the same throughput and fairness performance with the same feedback overhead yet with less computational complexity. Therefore, at such low loading levels, the quasi-FDR is more adequate provided that advances in technology would have created effective ways to resolve the quasi-FDR implementation challenges.

4.7 Conclusions

Significant throughput fairness and ubiquity can be achieved in a cellular relay network with symmetric inelastic traffic through formulating a throughput-optimal policy that performs joint routing and scheduling on frame-by-frame basis, e.g., the quasi-FDR scheme vs. the PFS. We present a novel throughput-optimal formulation in accordance with the emerging OFDMA-based cellular relay networks employing half-duplex relaying. Low-complexity iterative algorithms are devised to solve the formulated optimization over two consecutive sub-frames using the queue length coupling. Our numerical results show that with a slight complexity increase as compared to the quasi-FDR scheme, the network capacity for which the queues can be stabilized has been significantly increased, and hence fairness and ubiquity at high traffic loads, besides the substantial improvement in both latency and queue-awareness. The results also show that without empirical priority weights, our efficient implementation of throughput-optimal scheduling achieves a ubiquitous and fair service within each class of users (with symmetric traffic) and across classes of asymmetric traffic in a relative sense, on the time-average and long-term time scales. Load balancing among only active relays is jointly realized with the resource allocation.

for instance in number of fixed-size packets or fragments, could be interesting to examine.

Chapter 5

Integrating Self-Organizing Nomadic Relays into OFDMA Fixed-Relay Cellular Networks

5.1 Introduction

This chapter describes decentralized RRM methods devised for OFDMA-based multicellular networks featuring a high density of wireless relays of different characteristics and functionalities. We recall that wireless service providers of the next-generation cellular networks aim at achieving a ubiquitous very high-data-rate coverage in a *cost-efficient manner*. Therefore, saving the cost of the dense deployment of full-fledged BSs without incurring a backhaul cost, various forms of wireless relays could be envisioned to play different roles in these networks; such as extending coverage, increasing capacity, or operating in a cooperative manner to improve throughput and reliability [2]. Since the licensed band remains a scarce and an expensive asset, and given that a portion of the radio resources has to be invested in multihop relaying, enabling the bandwidth-intensive future services highlights the need for bandwidth-efficient relay-based RRM schemes [11]. Such relay-based RRM schemes should *cope with the system's dynamics*, exploit the characteristics of *different types of wireless relays*, and enable *aggressive* and *opportunistic* spatial resource reuse rather than adopting the traditional static reuse patterns. Nevertheless, in such scenarios with a high number

of network entities, decentralized RRM approaches have to be considered rather than the centralized approach which would result in a prohibitive computational complexity and feedback overhead.

Self-organization is also a highly desired feature by the service providers of next-generation networks (NGNs) towards dynamic, robust and scalable networks [113] featuring substantial savings in the operational expenditure (OPEX) and the capital expenditure (CAPEX) [114]. In self-organizing networks (SONs), some or all network entities operate and interact, either locally or in a distributed manner, to improve the overall network performance with the least human planning or administration and the least complexity and communication overhead [113]. Therefore, a self-organizing network entity possesses one function or more of self-configuration, self-optimization, or self-healing [114]. In fact, the high potentials in self-organization in terms of performance, robustness and cost savings have been driving the ongoing efforts in both research [115] and standardization [116] to incorporate this intelligence in the NGNs through various use cases. Nevertheless, due to the steadily growing interest in relay-enhanced cellular networks, managing wireless relays has received a considerable attention while laying out the framework for developing self-organizing methods in those use cases [117].

Therefore, considering only fixed relay stations (FRSs), some works have introduced self-organizing functionalities to the cellular network. In [118], for instance, an adhoc network of FRSs is overlaid on a conventional cellular architecture while the objective of the proposed algorithm is to design the topology of these FRSs to achieve the desired QoS, given a number of relays and a fixed coverage area. Whereas the proposed scheme in [119] allows the BS to select a receiving FRS at a given time slot of the single-carrier time division multiple access system, and then the BS transmits a preamble from which the other RSs know this allocation result and self-configure their status (i.e., receive, transmit, or shut down) based on each relay's set of close

neighbors. An interesting approach has been also discussed in [120] using FRSs deployed at the cell boundaries. The authors conceive that using smart antennas at the FRSs suffices to ensure high data rate backbone connections. The algorithms therein rely on passive direction of arrival estimation to coordinate the radiation patterns of the relays for load balancing and adapting coverage. A self-organizing load balancing framework has been proposed as well in [121] where each BS or FRS shares its own load measurements and resource information with a cluster of potential neighboring partners and then decides which load balancing policy, or combination of policies, to trigger.

However, it is envisaged that the NGN architecture will comprise a plethora of relay stations. Worthy of mention is the *plug-and-play* type of relay known as *nomadic relay*, an idea that has been entrenched in the IEEE 802.16 standards [18], [29], and gaining widespread acceptance. In fact, FRSs are privileged by their strategic geographical deployment with respect to the serving BS along with the off-the-wall power supply availability. While mobile relay stations (MRSs) are characterized by their mobility (e.g., rooftop-mounted vehicular devices), nomadic relay stations (NRSs) are technically stationary devices but portable, and like MRSs, are battery operated; much lower transmit power levels than those of FRSs are thus anticipated. Moreover, NRSs are **transparent to the operator's infrastructure and mostly deployed by the users**. It is clear though that the current literature on cellular relay networks does not distinguish between, and is thus unable to exploit, the different characteristics of these types of wireless relays.

Motivational scenarios for using multihop relaying are provided in the survey paper [54]. Therein, NRSs can be deployed to provide a temporary coverage and capacity in an area where FRSs may not provide the required QoS. Example of temporary coverage areas could be, in general, where wireless connectivity is required for only a short period of time, such as in trade fairs and sporting events or in

disaster recovery situations. In addition, NRSs can be used to spread the capacity in a large building. We also observe that an NRS -with much less constraints on size as opposed to a hand-held wireless terminal (WT)- may act as an extension to a close WT, providing a neat solution to the challenging problem of infeasible deployment of multiple antennas at hand-held terminals. Hence, enabling the benefits of (MIMO) techniques. In such scenarios, NRSs will coexist with FRSs yet with potentially much better communication links to the WTs.

This chapter therefore describes decentralized RRM methods featuring aggressive and opportunistic reuse in OFDMA-based multicellular networks enhanced with a mix of FRSs and self-organizing NRSs. To the extent of our knowledge, no work so far, other than our earlier contributions [122] and [123], has provided mechanisms for integrating the autonomous NRSs into the cellular network or suggested the underlying RRM schemes and protocols to facilitate their coexistence with FRSs. Our contribution is therefore pioneering and it represents an exposition and demonstration of feasible RRM schemes in facilitating the operation of nomadic relays. Since the current literature does not cover the systems and architecture considered in this chapter, we will refer to relevant works in the literature of decentralized RRM in OFDMA-based networks within our discussions.

In this work, we investigate the DL operation of a visionary wireless network model, where FRSs are augmented by NRSs with the aim of providing a more reliable cost-effective network. These NRSs will often be acting as intermediate nodes between a serving FRS and a WT forming a three-hop communication. To alleviate the burden on system resources due to multihop relaying, intelligent RRM schemes are needed to facilitate aggressive resource reuse and meanwhile combat the potential CCI through opportunistic reuse and interference avoidance mechanisms. The research contributions in this chapter can be summarized as follows:

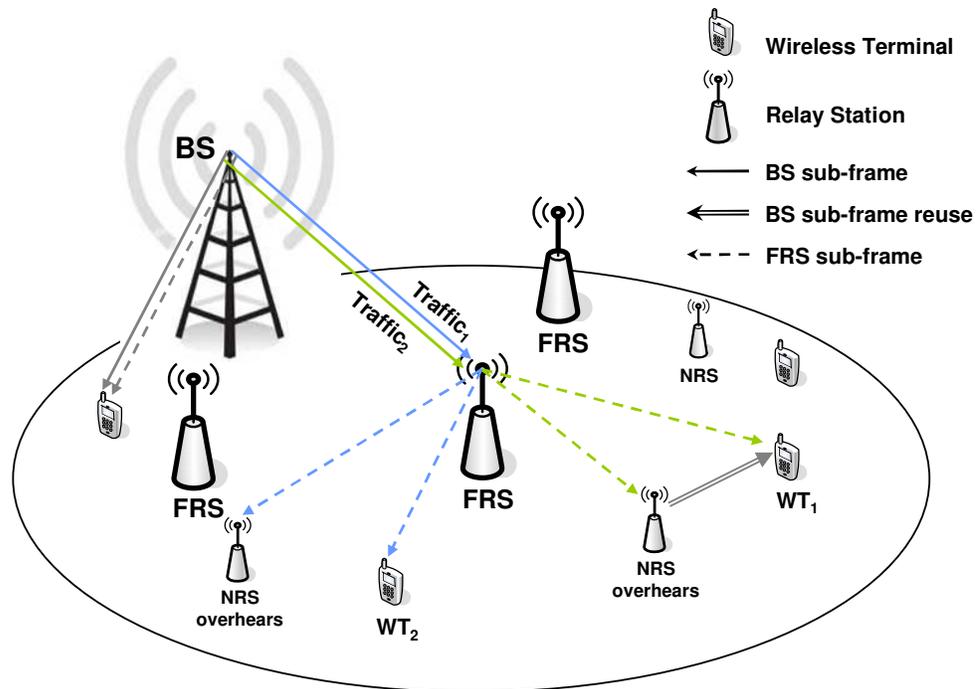


Figure 5.1: The region of interest may represent a cell of the hexagonal grid, or a sector of a cell, in the multicellular network. All resources are available in this region.

- We devise a self-optimizing user-based intra-cell routing (link selection) strategy that significantly reduces the amount of feedback overhead and achieves interference avoidance.
- We introduce self-optimizing opportunistic medium access (MAC) technique by which NRSs autonomously acquire radio resources. NRSs asynchronously sense the activities of the subchannels around the WT that needs assistance and autonomously acquire the subchannels that have least activity. This can succinctly be described as: Listen, Acquire Resources, then Assist (LARA). This technique can be imported into any other RRM scheme.
- Through the asynchronous opportunistic MAC of NRSs, a smart intra-cell reuse that exploits the channel and interference conditions is achieved in contrast to the commonly adopted static reuse patterns, e.g., [43].

- We devise a protocol to enable the cooperation between an NRS and an FRS to assist troubled WTs thereby boosting the performance of cell-edge users. The cooperation is based on selective relaying, meaning that the NRS only helps its WT on request and it cooperates only on the data segments it received reliably.
- We introduce novel limited-feedback heuristic decentralized RRA schemes to facilitate the operation of FRSs, the integration of NRSs, and mitigation of CCI in OFDMA-based multicellular networks.
- Finally, we conduct extensive performance evaluation of the proposed schemes and techniques under realistic channel models and system parameters.
- Thus, we establish the concept of nomadic relay-augmented fixed relay networks.

The rest of this chapter expounds on the above bullets. The journal paper [65], the patent fillings [124], the conference paper [122], as well as the technical reports [125] and [126] are the outcomes of the work presented in this chapter. Possible extensions and open research directions are provided in Chapter 7.

5.2 Description of the Proposed Systems

A representative region or ‘cell’ of the proposed system is shown in Fig. 5.1. It consists of a BS, FRSs, WTs, and NRSs. The same spectrum is available in each cell¹. The figure depicts the DL transmissions over two sub-frames and the half-duplex operation of all relay stations. The system shown may operate in either TDD or FDD. The basic resource allocation unit is the OFDM subchannel comprising a number of adjacent subcarriers; there are N data subchannels for each the channel fading is flat. AMC based on the channel quality is generally assumed. However, the NRS regenerates the

¹Without loss of generality, this cell could be an LTE-A ‘cell’ served by one of the three directional beams of an eNB [127].

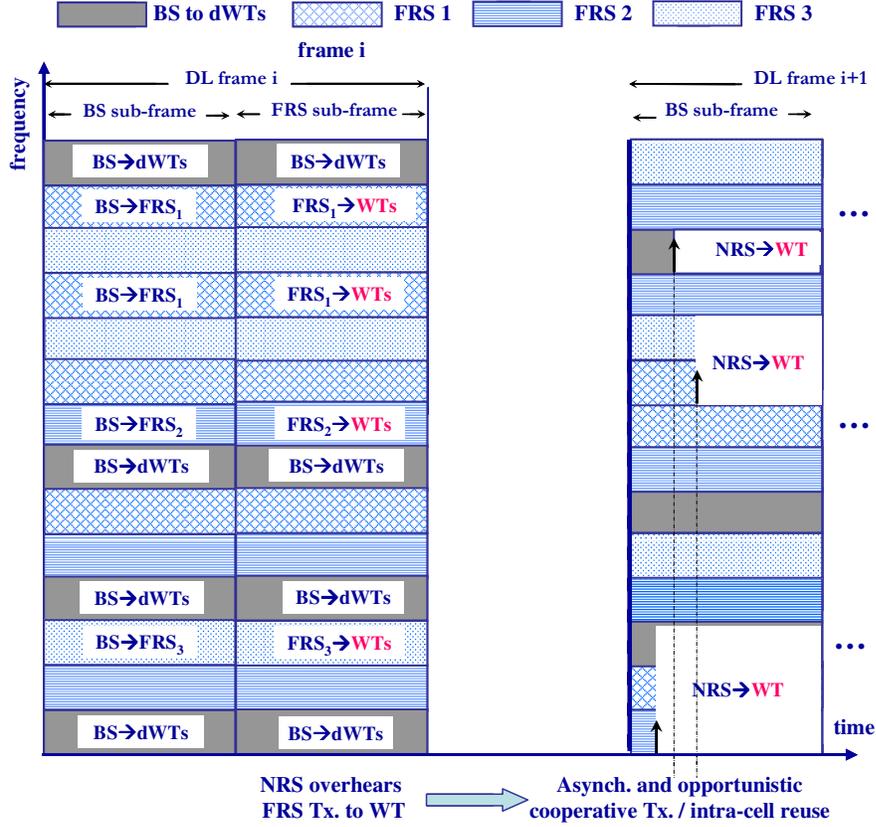


Figure 5.2: Frame structure of the proposed schemes showing 3 active FRSs and the overlaying of the NRS operation from a downlink frame to the following one after an uplink frame has elapsed.

received signal while preserving its modulation and coding levels rather than *adapting* them to the NRS-WT channel which requires a costly feedback from the WT. For simplicity, an upper-bounded version of the continuous adaptive modulation function in [28] is employed. Therefore, the achievable rate on subchannel n of the link from source m to destination k , at a particular target bit error rate P_e and subchannel bandwidth W , is an upper-bounded function of the received SINR $\gamma_{m,k,n}$, considering the CCI observed in the previous DL frame as in (5.1).

$$R_{m,k,n} = W \min \left\{ \log_2 \left(1 + \frac{-1.5\gamma_{m,k,n}}{\ln(5 P_e)} \right), 10 \right\} \text{ bps.} \quad (5.1)$$

The DL frame is partitioned into two consecutive equal-length sub-frames. An FRS receives only during the first sub-frame (BS sub-frame) and transmits during the

second (FRS sub-frame) as shown in the frame structure in Fig. 5.2. Note that the set of subchannels, \mathcal{N}_m , which is dynamically assigned to the feeder link of FRS_{*m*} is allocated among its connected WTs for the second hop of the same DL frame. As such, transmissions of different FRSs in the same cell are maintained orthogonal during the FRS sub-frame. Such an approach offers a compromise between the multiuser/frequency diversity gains and the CCI often realized in distributed schemes due to collisions of the uncoordinated allocation decisions of FRSs. A similar yet more stringent constraint has been imposed in [50] where the user data must be scheduled on the same subcarrier over the two hops. This is in contrast to our schemes where the subchannel pairing between the first and second hops is optimized by the RRA scheme in use.

There are M FRSs deployed in the cell at strategic locations where good line-of-sight (LOS) communication is maintained and highly directional antennas are employed making BS-FRS feeder links more immune to CCI with less shadowing and small-scale fading as compared to user access links. We therefore assume that the achievable rates on the feeder link of FRS_{*m*} are sufficiently good compared to those on its users' access links so that for any set of subchannels $\mathcal{N}_{m \rightarrow k} \subseteq \mathcal{N}_m$ assigned for the access of relayed user k during the FRS sub-frame, there exists another set of subchannels $\mathcal{N}_{0 \rightarrow m, k} \subseteq \mathcal{N}_m$ that can be used on the feeder link such that:

$$\sum_{j \in \mathcal{N}_{0 \rightarrow m, k}} R_{0, m, j} \geq \sum_{i \in \mathcal{N}_{m \rightarrow k}} R_{m, k, i}, \quad \forall m \neq 0, \quad (5.2)$$

where \mathcal{N}_m is the set of subchannels assigned by the BS (node 0) to FRS_{*m*} and $\bigcap_{m=0}^M \mathcal{N}_m = \phi$. This is further supported by the increasing frequency diversity on the feeder link as the number of relayed WTs by FRS_{*m*} increases along with the flexibility of the RRA scheme in subchannel pairing over the two hops. As such, the end-to-end capacity using equal sub-frames is governed only by the relayed user's

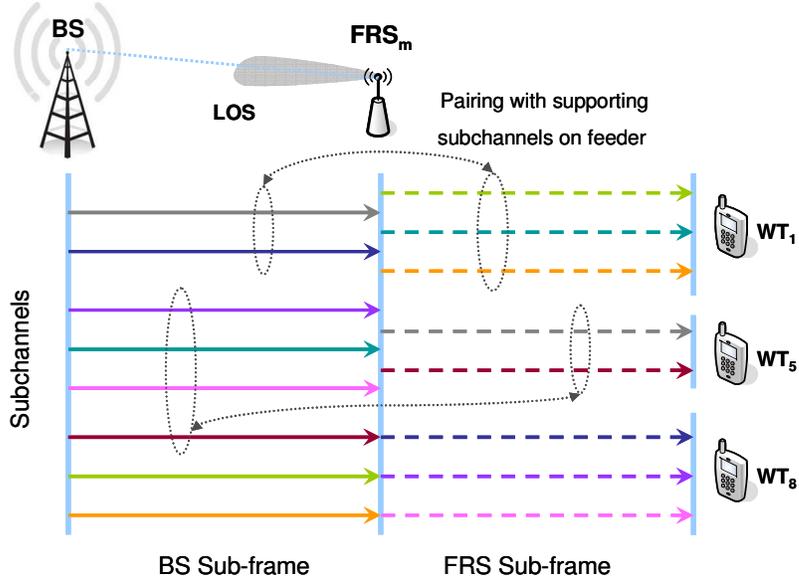


Figure 5.3: Examples of subchannel pairing between the feeder link of FRS_m and the access links of its connected WTs.

achievable rates on the second hop; this is inline as well with the decentralized nature of the NGNs where a WT can measure the quality of its access links but may not know the channel states of each potential feeder link to self-optimize its routing decision, for instance, instead of the BS. Figure 5.3 shows examples of subchannel pairing between the feeder link of FRS_m and the access links of its connected WTs. Note that the same set of subchannels is used over the two sub-frames. While the FRS_m - WT_1 access link has been assigned, for instance, 3 subchannels in this figure, the feeder link could support the total achievable rate of this link using only 2 subchannels (not necessarily in $\mathcal{N}_{m \rightarrow 1}$). On the other hand, 3 subchannels have been paired with those of the FRS_m - WT_8 access link with an extra capacity that has been used by the BS to support the total rate on the FRS_m - WT_5 access link which happened to be of high instantaneous quality.

Note that the deployment of FRSs could be different from one cell to another since our schemes are not attached to a certain geography. The WT dynamically selects one access link out of $M + 1$ (either through the BS or an FRS) rather than the static relay selection commonly adopted in literature, e.g., [50], and [48]. There are K active WTs in the cell and K_{nom} NRSs. Depending on the scenario of interest, the number

of NRSs may vary with respect to K . For the sake of illustration, we have considered the case in which $K_{nom} = K$; that reflects on a scenario where each WT has placed, or selected out of many, an NRS that is dedicated to assist it on demand, e.g., an indoor NRS that is attached to the window. The NRS is thus stationary, in a close vicinity of the WT, and often has a very good connection to the WT; likely an LOS. The NRS is conceived to be larger than a WT and thus can employ more antennas to strengthen its link to either the serving FRS or the WT through beam steering. Note that the potential serving FRSs are stationary, while the NRS can be equipped with the computational power for its beam to follow a mobile WT. That is because NRSs do not handle upper layer issues such as applications and human interfaces. Fixed power (per subchannel) allocation is assumed for BSs and relays. The transmit power per subchannel of the BS is greater than that of an FRS and much greater than that of an NRS. The choice of low transmit power for NRS is informed by its proximity to the WT so that NRSs do not unnecessarily cause interference to other links. Continuous backlog is assumed at the BS buffers but not at the FRSs. In general, the basic operation of the schemes can be described as follows.

In the BS sub-frame, the BS transmits to the direct WTs (dWTs) and the FRSs with connected WTs while any NRS can access one or more subchannels at a random instant (immediately after listening to the activity on all subchannels), and transmit to a particular WT during the remaining time of the same sub-frame. The NRS's opportunistic medium access results in intra-cell (or intra-region) reuse if an acquired subchannel is allocated to a BS-FRS feeder link or a BS-dWT access link. In the FRS sub-frame, the BS continues to communicate with the dWTs on the same set of subchannels assigned to them in the previous sub-frame; note that the channel quality of all links is invariant over the DL frame duration. FRSs transmit to their connected WTs while the NRSs overhear and always decode the transmissions destined to their respective WTs.

Wireless terminals can receive from multiple nodes simultaneously on orthogonal subchannels. This happens if a WT is connected directly to the BS while receiving from its NRS. Note that due to the half-duplex operation, an NRS may not overhear and transmit concurrently. For the sake of illustration and emphasizing the concept, we consider a protocol where NRS cooperates with only FRSs². Therefore, the NRS transmits during the BS sub-frame selected data segments out of the data received from the FRS during the past DL frame. This selective-relaying cooperative protocol is different from an automatic repeat request (ARQ) through the serving FRS for the following reasons:

1. Latency is significantly less since the NRS reacts after one UL frame duration (in the new BS sub-frame) whereas the serving FRS can only retransmit in the following FRS sub-frame
2. The FRS would have to utilize a portion of the system's premium resources to retransmit whereas the NRS cooperates through intra-cell reuse.
3. Given the network geometry and physical dimensions, a WT and its NRS have almost the same distance from both the serving and interfering FRSs in the network, whereas most realistic spatial shadowing correlation models result in the same shadowing realization for the FRS-WT and its associated FRS-NRS links [128]. As such, this protocol may exploit the spatial diversity offered by the FRS-NRS link which enjoys a better link budget through the NRS directional antenna gain.
4. The NRS in this protocol does not change the AMC level of the received signal relying on its proximity to the WT and therefore significantly saves in the

²Based on the WT-based dynamic routing, different protocols can be devised by imposing constraints on either the routing options or the NRS assistance to facilitate its cooperation with the serving BS, serving FRS, or both.

feedback and control signalling which would have been required to update the FRS with the new channel state information (CSI) and to set up the WT receiver to the new AMC, respectively.

Having described the system and the generic framework of the proposed RRM schemes, in the following section we discuss in details the first self-organizing component which is hosted by the WT.

5.3 The Self-Optimizing WT-based Dynamic Routing

We refer to the proposed intra-cell routing strategy as WT-based routing or link selection. Each WT selects its serving node for the following DL frame, i.e., a WT can be served directly by the BS or through an FRS. The challenge here is how the WT compares the radio access links while the quality of each varies over N subchannels. A key fact is that a WT will end up being scheduled on only a subset of the N subchannels on the selected link; such subset potentially encompasses the link's best subchannels rather than the poor ones. As such, our proposed self-optimizing routing strategy can be implemented as follows and illustrated in Fig. 5.4:

1. During DL frame i , the k -th WT assesses the quality of the subchannels on the $M + 1$ access links, then sorts each link subchannels in a descending order of their achievable rates.
2. WT selects a certain percentile of the best subchannels on each link.
3. A statistical metric of the selected subchannels is computed, e.g., the mean value.
4. WT selects the access link, from node m^* , with the largest metric.

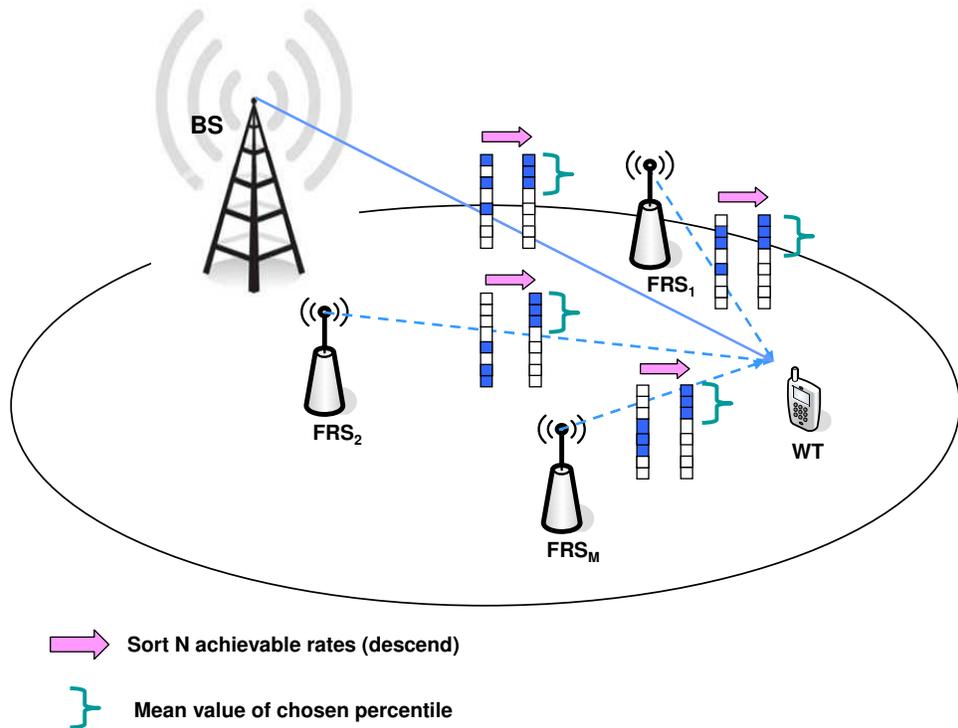


Figure 5.4: The WT-based self-organizing routing or link selection strategy.

5. WT sends the CSI feedback vector $\mathbf{r}_{m^*,k} = [R_{m^*,k,1}, \dots, R_{m^*,k,N}]^\top$ to only the selected node (an FRS or the BS) during the uplink frame i .³ Hence, the routing decision is known to the concerned node and implicitly to all other nodes with substantial savings in feedback overhead.

There is a tremendous advantage in this routing strategy compared to the conventional approach, e.g., [8], where the subchannels on each potential serving link are reported to the BS. Note that further savings can be achieved by replacing the continuous achievable rates by indexes of the discrete AMC modes when AMC lookup tables are used in practise.

We define $\bar{R}_{m \rightarrow k}^{(i)}$ as the average achievable rate of the selected percentile on the user's access link from node m during the DL frame i . The k -th WT's route selection

³Notation: Lowercase bold face symbols such as \mathbf{r} denote vectors, uppercase bold face symbols such as \mathbf{R} denote matrices, whereas uppercase calligraphy such as \mathcal{U} denote sets of elements.

is then based on the following criterion:

$$\mathcal{K}_{m^*}^{(i+1)} \leftarrow \mathcal{K}_{m^*}^{(i+1)} \cup \{k\}, \quad m^* = \arg \max_m \{\bar{R}_{m \rightarrow k}^{(i)}\}, \quad \forall m, \quad (5.3)$$

where $\mathcal{K}_m^{(i+1)}$ is the set of WTs connected to node m during DL frame $i + 1$.⁴ We note that this self-optimizing routing strategy remains the same under the two decentralized RRA heuristics discussed in Section 5.5.

A WT-based link selection (or transmission mode selection) criteria has been also proposed in [129] for the WT to choose between the BS and a single FRS in its sector based on a utility function that is equivalent to the rate utility per resource unit price. The rate utility relies however on the end-to-end average rate as a function of the average SINRs of the access links and the feeder link. Interestingly, the authors have suggested that the WT ‘irrationally’ violates the utility criteria with a preset probability to avoid local optimal decisions.

In the following section we discuss the functionality and the self-organizing aspects of NRSs as integrated into our FRS-based network.

5.4 Self-Organizing Nomadic Relay Operations

Figure 5.5 demonstrates the operations of the NRS in resource acquisition and cooperation with an FRS to assist a troubled WT. This protocol can succinctly be described as: Listen, Acquire Resources, then Assist (LARA) and can be imported into any other RRM scheme in multicarrier systems.

During the FRS sub-frame, the FRS transmits to its connected WTs on the allocated subchannels by the RRA scheme while the BS continues to transmit to the dWTs. An NRS dedicated to (or paired with) a particular WT always overhears and

⁴The frequency of executing this strategy can be generally relaxed in time so that the routing decision is changed after averaging the metric over a window of several frames.

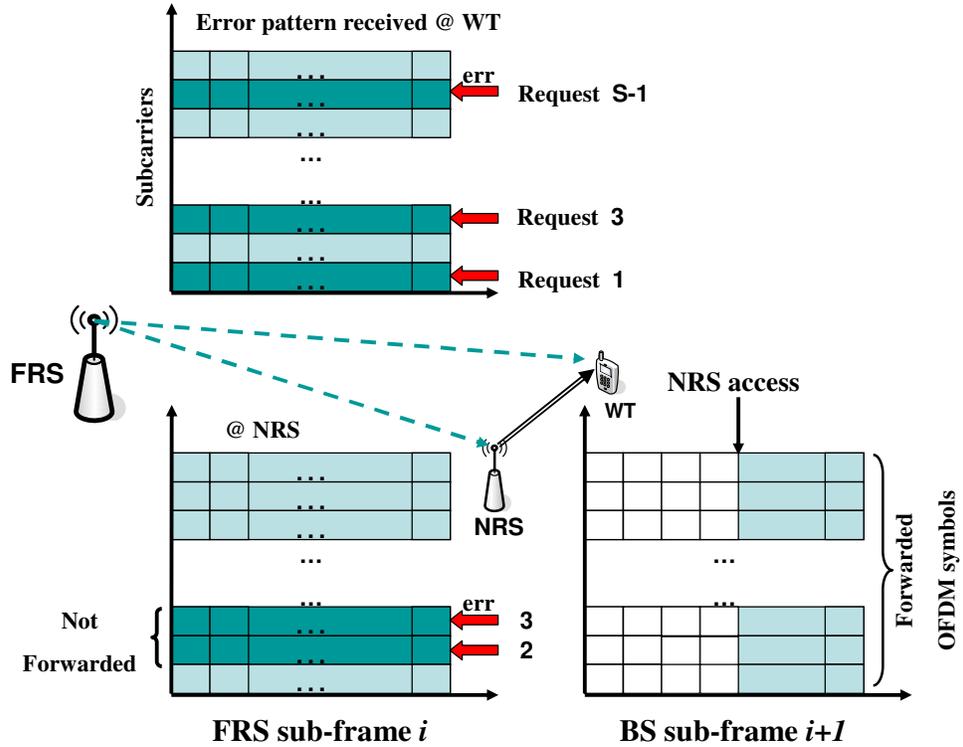


Figure 5.5: Illustration of the cooperation of the serving FRS and the dedicated NRS to assist the WT. S is the number of OFDM subcarriers per subchannel.

reliably decodes the transmissions from the serving FRS. In the next BS sub-frame, the BS transmits to some dWTs and FRSs while the NRS can access one or more subchannels to forward the data it has *reliably decoded* to the WT spontaneously or based on the WT's request upon reception of erroneous data segments⁵. Generally, traditional combining techniques (selection or maximal ratio combining) may be invoked at the WT side thereafter utilizing the fact that the NRS preserves the AMC modes of the original FRS transmission.

5.4.1 Autonomous Subchannels Acquisition and Selective Relaying

In the proposed scheme the NRS autonomously chooses the subchannel(s) it needs to communicate with the WT. The NRS chooses a random instant in the interval $[\epsilon, T_{end}]$ to listen to the whole bandwidth and estimates, without detection, the amount of total

⁵Refer to Appendix B for channel emulation and generation of error patterns at the WTs and NRSs with lossless NRS-WT links.

received power on each subchannel (see Fig. 5.5). The parameters ϵ and t_{end} can be set by the service provider to represent respectively the minimum time required for the hardware to start listening from the initiation of the BS sub-frame (of duration T_1) and the minimum time allowed for NRS transmission, i.e., $t_{end} \leq T_1 - T_{min}$. Based on the proximity of the NRS to the WT, the total power received at the NRS on a subchannel during this short listening period can be a good approximation of the interference level that could be observed by the WT if the NRS uses that subchannel for the remaining sub-frame duration. The NRS will examine all the N subchannels except those assigned to the WT (if connected directly to the BS in that frame). Such exception prevents a situation where the WT is supposed to receive from both the BS and the NRS on the same subchannel.

Due to the asynchronous access in the continuum of time $[\epsilon, T_{end}]$, the probability of having more than one NRS listening at the same instant, and furthermore, interfering significantly with one another's transmission, is negligible [130]. The NRS sorts the observed subchannels based on the estimated total received power and acquires the subchannel(s) with the least power. The number of chosen subchannels can be determined if we consider that the NRS attempts to minimize latency by forwarding all the data segments it received properly, and the WT requires, during the remaining BS sub-frame duration and using the same AMC modes (see Fig. 5.2).

Let Γ denote the number of subcarriers to forward due to one subchannel detection, out of S subcarriers per subchannel, while s denotes the number of OFDM symbols per one subcarrier. At the time of NRS channel access, let us denote the remaining number of symbols per subchannel (the shaded part in Fig. 5.5) as ω . Therefore, the number of OFDM symbols to forward is given as

$$N_f = \Gamma \times s, \tag{5.4}$$

while the total number of required (reuse) subchannels to assist the WT with the detection of that single subchannel is given as

$$n_r = \left\lceil \frac{N_f}{\omega} \right\rceil. \quad (5.5)$$

The key idea behind the asynchronous access is to allow the NRS to sense the activity of other NRSs that are close in distance, or more precisely, with high interfering-link gains; thus minimizing the likelihood of choosing the same subchannels immediately acquired by such NRSs. A similar idea to our asynchronous listening and updating has been employed for ad-hoc networks in [130] where they rely strongly on channel reciprocity to allow a cluster head to estimate the interference it will cause to the others when it uses a particular band. In contrast, this condition is not necessary in our schemes since the NRS only listens to estimate the interference that its WT will experience on a subchannel if the NRS uses it to serve the WT. It is worth mentioning that LARA, being a *non-contention based* protocol, is different from the carrier-sense multiple-access/collision avoidance (CSMA/CA) technique used in 802.11 standards. LARA operates in OFDMA-based cellular system where excessive CCI is avoided through acquiring the carriers with the least activity observed. A subchannel could be *opportunistically acquired by several NRSs without a back-off delay* mechanism, given the proximity of the NRS to its WT and the low NRS transmit power. Nevertheless, LARA is invoked on demand (not all overhearing NRSs will attempt to access the medium), handles selected segments of the overheard data, and its medium access is limited to the interval $[\epsilon, t_{end}]$.

More importantly, in underloaded network conditions where the system resources are not fully utilized, such MAC technique gives the NRSs the opportunity to *first acquire the unoccupied resources* before reusing the occupied ones (by the BSs in that scenario). In contrast, any static spatial reuse would unnecessarily result in excessive CCI disregarding the unoccupied resources.

5.4.2 NRS-FRS Cooperation

While the ‘data segment’ on which the NRS selective relaying operates may refer to any data unit such as a packet or a subchannel payload, we here consider, without loss of generality, a subcarrier as the data segment as shown in Fig. 5.5. By selective relaying, the NRS only requires a fraction of the resource used by the FRS to forward its data; this is particularly useful since the NRS only accesses the channel over a fraction of the BS sub-frame.

It is worth noting that the NRS-FRS cooperation within the LARA protocol is in principle similar to the conventional digital cooperative relaying protocols where the relay overhears the source-to-destination transmission in the first time slot and then regenerates and forwards the decoded message to the destination in the second slot. However, LARA using its overlay from frame i to $i + 1$, takes advantage of the cellular frame structure where an uplink frame always occurs between the first and the second time slots (sub-frames). In that intermediate uplink frame, the WT has an opportunity to inform the NRS of which data segments to forward, and thus the heavy burden on the system resources is alleviated when few, or sometimes no, subchannels are acquired. To the best of our knowledge, such an advantage is unprecedented as the cooperative relaying schemes available in the literature do not have this feature.

5.5 Radio Resource Allocation at the BS

The BS’s RRA algorithms for both decentralized schemes, namely the ‘distributed’ and the ‘semi-centralized’, operate in a greedy manner to maximize the overall system throughput under the users’ QoS constraints, the cell-edge users inclusive. The QoS requirements are represented by the user’s minimum rate R_{min} and the target BER P_e . Based on the earlier assumptions and system model, the BS can allocate the resources dynamically among all the dWTs and the feeder links of the FRSs based on their second hops, as a two-dimensional assignment problem.

Therefore, the Hungarian algorithm [131] is used to provide a low-complexity iterative solution in which each iteration is solved optimally incurring a polynomial complexity of $\mathcal{O}((\max\{|\mathcal{U}|, |\mathcal{K}_0| + |\mathcal{M}_a|\})^3)$ in the distributed scheme, and $\mathcal{O}((\max\{|\mathcal{U}|, K\})^3)$ in the semi-centralized, where $|\mathcal{U}|$ and $|\mathcal{M}_a|$ are respectively the number of unassigned subchannels and the number of active feeder links, i.e., with connected WTs. The matrix passed to the Hungarian algorithm is constructed using the achievable rates feedback by the FRSs and the dWTs as shown in Fig. 5.6 and Fig. 5.7.

After each iteration, the minimum rate requirement of each column is compared to its assigned sum rate. When the rate requirement is satisfied for a column, it is excluded from the subsequent iterations. As such, the final assignments are not necessarily equal in the total number of subchannels.

Since the dWTs, unlike the relayed WTs, remain connected for the whole DL frame, the minimum rate requirement of a relayed WT is seen as double that of a dWT, given the equal sub-frames⁶. If all columns are rate satisfied before the subchannels are exhausted, the operator has the flexibility to assign the remaining subchannels among the dWTs and/or the feeder links that have the highest achievable rates. To put more emphasis on the impact of NRSs, given the FRS-NRS cooperation protocol we adopted, the remaining subchannels are allocated in a greedy manner among the feeder links only, and thus among the users relayed through the FRSs.

5.5.1 Semi-Centralized Scheme

Here, the BS performs the RRA for the relayed WTs and the dWTs without the FRSs taking part in the optimization. However, as explained in Section 6.2, the scheme still has limited-feedback, NRSs operate autonomously, and the routing is carried out by the WTs. Hence, the scheme is described as semi-centralized. The achievable rates of the relayed WTs at FRS_{*m*} are arranged in an $N \times |\mathcal{K}_m|$ matrix and conveyed to

⁶In the generic case of unequal sub-frames (e.g., T_1 and T_2 , respectively), the minimum rate requirement of a relayed WT is seen as αR_{min} where $\alpha = \frac{T_1 + T_2}{T_2}$.

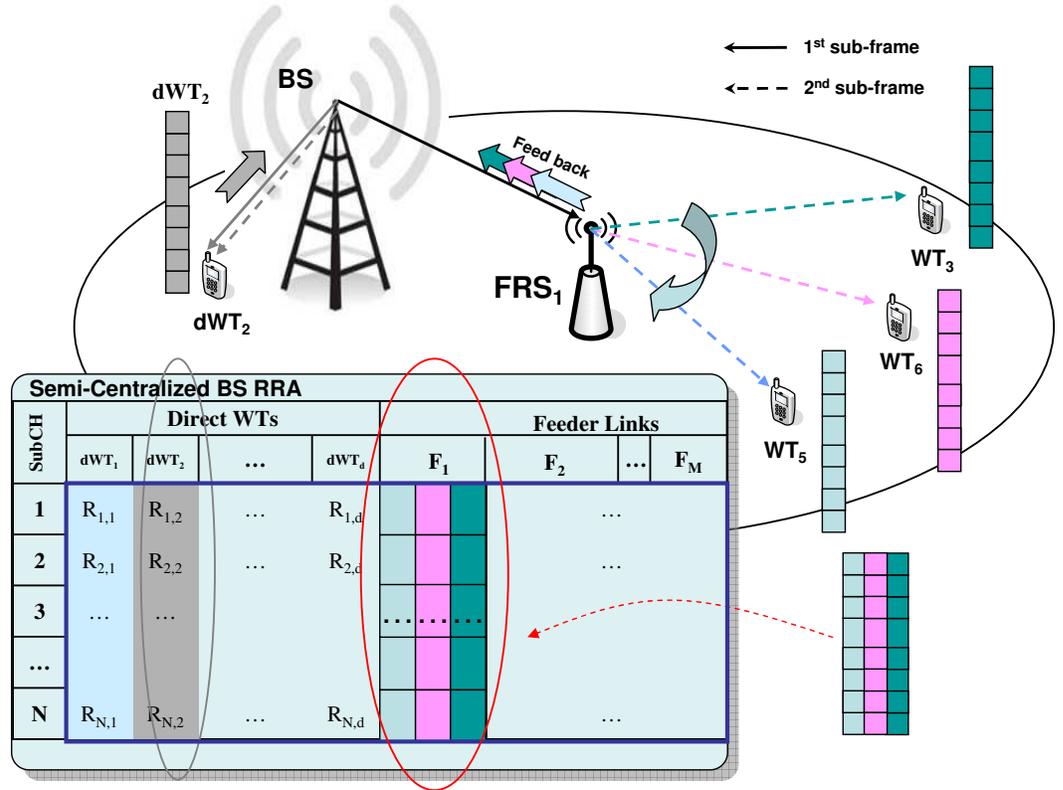


Figure 5.6: Illustration of the operation of the heuristic semi-centralized RRA scheme. the BS during the UL frame to form an $N \times K$ matrix as shown in Fig. 5.6.

5.5.2 Distributed Scheme

During the UL frame, each FRS appears as a large WT by masking its connected WTs, \mathcal{K}_m , and combining their channel state information (CSI) feedback into a single vector through a different implementation of the Hungarian algorithm regardless of the minimum rate constraints. As shown in Fig. 5.7, this vector is intended to be evenly shared by the actual rates and therefore can be considered a wish-list that provides unbiased representation of the second hops altogether; this vector could be influenced, under a different processing, by the worst or the best relayed user's link. In a related work [50], such processing is performed at the FRS using the average or median, per subchannel, of the SINRs across all the connected users. As such, the reported values therein do not reflect on any user's actual subchannel quality. Moreover, an overlooked problem arises from the BS's perspective: How many data bits should the BS send to each relayed user given the processed CSI?

We provide a neat solution to such problem; the index of the user whose subchannel n has been selected while forming the combined CSI vector, is recorded in the n^{th} entry of an auxiliary vector, i.e., the FRS passes the final information to the BS in two $N \times 1$ vectors. The BS performs the optimization and then estimates the number of information bits to be forwarded for each relayed WT based on the user index corresponding to the assigned subchannel from the combined column. We also note that when the feeder link appears as a single processed column in the BS RRA matrix, the feeder will have a virtual minimum rate requirement that is equal to double the sum of its connected users' rate requirements.

The RRA at the FRSs: Due to the combined feedback of the relayed WTs under the distributed scheme, the BS RRA does not guarantee that the rate requirement of each user will be met. Therefore, the FRS starts a separate re-allocation process on the set of subchannels that has been assigned to its feeder link \mathcal{N}_m . The FRS performs a low-complexity optimization process similar to the BS's but on an $|\mathcal{N}_m| \times |\mathcal{K}_m|$ rate matrix. The rate constraints are checked after each assignment. Remaining subchannels after satisfying the constraints are assigned to the best users.

In the following (see charts below) we present the pseudo-codes for the semi-centralized and the distributed RRA heuristic algorithms at the BS. In these codes, \mathcal{U} , \mathcal{N} , \mathcal{K} , \mathcal{M}_a , and \mathcal{C} denote the sets of unassigned subchannels, all available subchannels, all WTs, active FRSs, and working columns in the achievable rate matrix \mathbf{R} , respectively. Note that $\mathbf{e}_L = [1 \dots 1]_{1 \times L}$.

5.5.3 Required Feedback Overhead

We now address the *feedback* required to realize the proposed schemes. The CSI feedback in the form of per-subchannel achievable rate is made available at the transmitting nodes as follows;

- BS-dWTs: A dWT needs to feed back its CSI to the BS; an $N \times 1$ vector.

Pseudo-code for the semi-centralized RRA algorithm at the BS

1. Initialization: Set $\mathcal{N}_m = \emptyset \forall m, \mathcal{U} = \mathcal{N}, \mathcal{C} = \mathcal{K}$.
 2. **for** each $c \in \mathcal{C}$
 3. $\mathbf{R} \leftarrow [\mathbf{R}|\mathbf{r}^c]$, % $\mathbf{r}^c = \mathbf{r}_{m^*,k}$ from routing results
 4. **if** $m^* \neq 0$ then
 5. $R_{min}^c = R_{min}$
 6. **else**
 7. $R_{min}^c = 2R_{min}$
 8. **end if**
 9. **end for**

 10. **while** $\mathcal{U} \neq \emptyset$ and $\mathcal{C} \neq \emptyset$ **do**
 11. $(\hat{\mathbf{n}}, \hat{\mathbf{c}}) \leftarrow \text{Hungarian}(-\mathbf{R})$ % Default is Min.
 12. % $\hat{\mathbf{n}}, \hat{\mathbf{c}}$ outputs are vectors of indices
 13. $\mathcal{U} \leftarrow \mathcal{U} - \{\hat{\mathbf{n}}\}$, $N_{assigned} = |\hat{\mathbf{n}}| = |\hat{\mathbf{c}}|$
 14. % $N_{assigned} \leq \min\{|\mathcal{U}|, |\mathcal{C}|\}$
 15. **for** $i = 1$ to $N_{assigned}$
 16. $\hat{n} \leftarrow \hat{\mathbf{n}}(i)$, $\hat{c} \leftarrow \hat{\mathbf{c}}(i)$
 17. $X^{\hat{c}} \leftarrow X^{\hat{c}} + \mathbf{r}^{\hat{c}}(\hat{n})$, $\mathcal{N}_{\hat{m}^*} \leftarrow \mathcal{N}_{\hat{m}^*} \cup \{\hat{n}\}$
 18. $\mathbf{R}(\hat{n}, :) = -\infty \mathbf{e}_K$ % Disable assigned subchannel row
 19. **if** $X^{\hat{c}} \geq R_{min}^{\hat{c}}$ then
 20. $\mathcal{C} \leftarrow \mathcal{C} - \{\hat{c}\}$
 21. $\mathbf{R}(:, \hat{c}) = -\infty \mathbf{e}_N^\top$ % Disable satisfied column
 22. **end if**
 23. **end for**
 24. **end while**
-
-

Pseudo-code for the distributed RRA algorithm at the BS

1. Initialization: Set $c = 0$, $\mathcal{N}_m = \emptyset \forall m$, $\mathcal{N}_m = \emptyset \forall m$, $\mathcal{U} = \mathcal{N}$,
 2. load $\mathcal{K}_m \forall m$, $\mathcal{C} = \mathcal{K}_0 \cup \mathcal{M}_a$.
 3. **for** each $k \in \mathcal{K}_0$
 4. $R_{min}^{c+1} = R_{min}$
 5. $\mathbf{R} \leftarrow [\mathbf{R}|\mathbf{r}^c]$, % $\mathbf{r}^{c+1} = \mathbf{r}_{0,k}$ from routing results
 6. **end for**
 7. **for** each $m \in \mathcal{M}_a$
 8. $R_{min}^{c+1} = 2|\mathcal{K}_m|R_{min}$
 9. $\mathbf{R} \leftarrow [\mathbf{R}|\mathbf{r}^c]$, % $\mathbf{r}^{c+1} = \mathbf{r}_{F_m}$ from combined feedback at FRS_m
 10. **end for**

 11. **while** $\mathcal{U} \neq \emptyset$ and $\mathcal{C} \neq \emptyset$ **do**
 12. $(\hat{\mathbf{n}}, \hat{\mathbf{c}}) \leftarrow \text{Hungarian}(-\mathbf{R})$ % Default is Min.
 13. % $\hat{\mathbf{n}}, \hat{\mathbf{c}}$ outputs are vectors of indices
 14. $\mathcal{U} \leftarrow \mathcal{U} - \{\hat{\mathbf{n}}\}$, $N_{assigned} = |\hat{\mathbf{n}}| = |\hat{\mathbf{c}}|$
 15. % $N_{assigned} \leq \min\{|\mathcal{U}|, |\mathcal{C}|\}$
 16. **for** $i = 1$ to $N_{assigned}$
 17. $\hat{n} \leftarrow \hat{\mathbf{n}}(i)$, $\hat{c} \leftarrow \hat{\mathbf{c}}(i)$
 18. $X^{\hat{c}} \leftarrow X^{\hat{c}} + \mathbf{r}^{\hat{c}}(\hat{n})$
 19. $\hat{c} \mapsto \hat{m}$, $\mathcal{N}_{\hat{m}} \leftarrow \mathcal{N}_{\hat{m}} \cup \{\hat{n}\}$
 20. $\mathbf{R}(\hat{n}, :) = -\infty \mathbf{e}_{|\mathcal{K}_0 \cup \mathcal{M}_a|}$ % Disable assigned subchannel row
 21. **if** $X^{\hat{c}} \geq R_{min}^{\hat{c}}$ **then**
 22. $\mathcal{C} \leftarrow \mathcal{C} - \{\hat{c}\}$
 23. $\mathbf{R}(:, \hat{c}) = -\infty \mathbf{e}_N^\top$ % Disable satisfied column
 24. **end if**
 25. **end for**
 26. **end while**
-
-

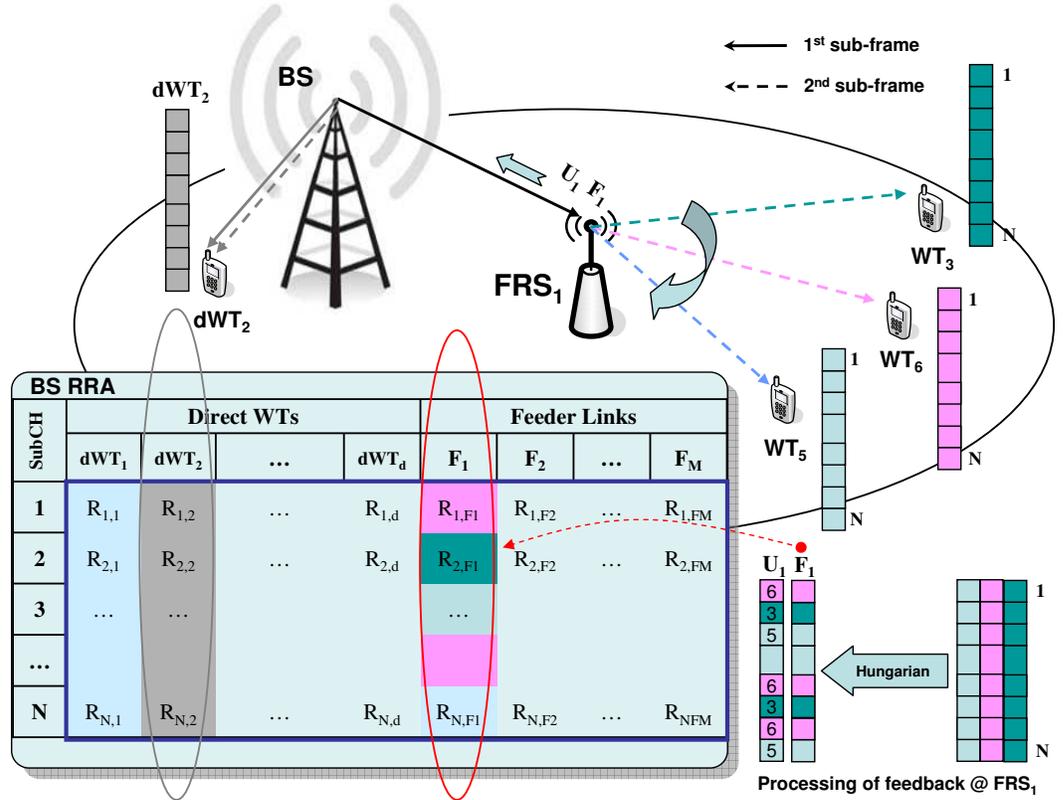


Figure 5.7: Illustration of the operation of the distributed RRA scheme and the achievable rates matrix at the BS.

- FRS-WTs: A relayed WT needs to feed back its CSI to only the selected serving FRS (an $N \times 1$ vector).
- BS-FRSs: In addition to the vector reporting the feeder link, an FRS reports the second hop depending on the RRA scheme and the result of WT-based dynamic routing:
 - Semi-centralized: A matrix of dynamic size $N \times |\mathcal{K}_m|$ from the FRS representing the subchannels' states of its connected WTs; if $\mathcal{K}_m = \emptyset$ in the upcoming frame, no feedback is sent.
 - Distributed: Two $N \times 1$ vectors representing the processed output of the FRS, as explained in Section 5.5.2. See Fig. 5.7.
- NRS-WT: No CSI feedback is required from the WT.

5.6 Simulation Results

Matlab system-level simulations have been conducted. The same QoS requirements (target minimum rate and maximum BER) are assumed for all WTs. The minimum rate requirement is set to 250 Kbps and the target BER is 10^{-3} . The 10th percentile is used in the WT-based routing. The simulated cellular network consists of 19 hexagonal cells enhanced with 3 or 6 FRSs. These relays are placed at a distance of 0.65 of the cell radius from the BS and with a uniform angular spacing. The distance between two adjacent BSs is 2 Km. Users are uniformly distributed within the cell area. It is assumed that the NRS is placed randomly within a distance of at most 20 m from the WT.

Time-frequency correlated small-scale fading is assumed; Rayleigh for NLOS links and Rician for LOS links. Independent lognormal shadowing is considered for different links in the network. As for the FRS-WT link and the corresponding FRS-NRS link, the same shadowing realization is applied to both links as different shadowing spatial correlation models result in an almost unity correlation coefficient. The NRS transmit/receive smart antenna gain is set to 7 dB. The path-loss model for the links as a function of the distance (in meters) is given as $PL = 38.4 + 35 \log_{10}(d)$ dB. Each FRS has an omni-directional transmit antenna to communicate with the WTs as well as a highly directive receive antenna aiming at the BS with LOS communication. Based on our system model, the quality of feeder links can be sufficiently higher than second hops to WTs and therefore they are not simulated to reduce the computational burden in the already complicated scenario.

A list of channel and system parameters used for the simulations are given in Table 5.1. Most of the parameters are taken from the 3GPP LTE release 9 (Case 3) [112] or the WiMax Forum [105] while the WINNER C2 channel model [92] is used.

Table 5.1: System parameters

Parameter	Value
User min. close-in distance to BS	35 m
BS Tx. antenna gain	15 dB
FRS Tx. antenna gain	10 dB
NRS Tx. & Rx. antenna gain	7 dB
WT Rx. antenna gain	0 dB
Shadowing std. dev. on user and interference links (NLOS)	8.9 dB
Shadowing std. dev. on FRS-NRS links (NLOS)	7 dB
Shadowing std. dev. on NRS-WT links (LOS)	1.5 dB
Carrier frequency	2.5 GHz
Total bandwidth	20 MHz
User mobility	10 Km/hr
Channel sampling time = TDD frame length	5 msec
Downlink : Uplink ratio	2:1
DL Tx. time in OFDM data symbols	24 symbols
NRS min. Tx. time T_{min} in OFDM data symbols	6 symbols
OFDM subcarrier bandwidth	10.9375 KHz
OFDM symbol duration	102.86 μ sec
Subchannel width	18 subcarriers
Noise power density at Rx. nodes	-174 dBm/Hz
BS total Tx. power	46 dBm
RS total Tx. power	37 dBm
NRS total Tx. power	19 dBm

The time-averaged and instantaneous user throughput: The CDFs of the time-averaged throughput in Fig. 5.8 show the throughput gains under both RRA schemes, due to the proposed NRS cooperation to assist the relayed WTs. Whereas the cell-edge performance is demonstrated by the lower-tail behavior of the CDFs of the time-averaged user throughput as shown in Fig. 5.9 considering the semi-centralized scheme with and without NRSs, and $K = 15$. According to the LTE evaluation methodology, the 5th percentile throughput corresponds to the cell-edge. Note that the time average is calculated for each user in each drop of given shadowing, WT location and NRS location. It can be observed that given the same K and M , the semi-centralized scheme outperforms the distributed in general and at the cell-edges in particular. This is due to the fact that the BS directly optimizes the RRA of the relayed users, who likely include the cell-edge users as a result of the WT-based routing, using their individual feedback information rather than allocating the resources to the serving FRSs based on combined feedback and QoS requirements. Given the rate matrix composition, i.e., $K \geq |\mathcal{K}_0| + |\mathcal{M}_a|$, the semi-centralized scheme better exploits the multi-user and spatial diversities than the distributed.

The performance gain due to the NRS assistance is clearly evident in both schemes. However, due to the considered uniform distribution of FRSs and WTs, along with the WT-based routing strategy which is common in both schemes, the number of relayed users increases as the number of FRSs increases. That is because a WT is more likely to find a neighbor FRS with a good radio access link quality under a higher FRS density of deployment. Since the NRS cooperation we considered is limited to the relayed users, the performance gain due to the employment of NRSs is amplified as the number of FRSs increases.

The same relative performances are realized through the outage probabilities in Table 5.2 based on the user's instantaneous throughput rate R_k achieved in each DL frame, as compared to the minimum required rate, i.e., $\mathcal{P}\{R_k < R_{min}\}$.

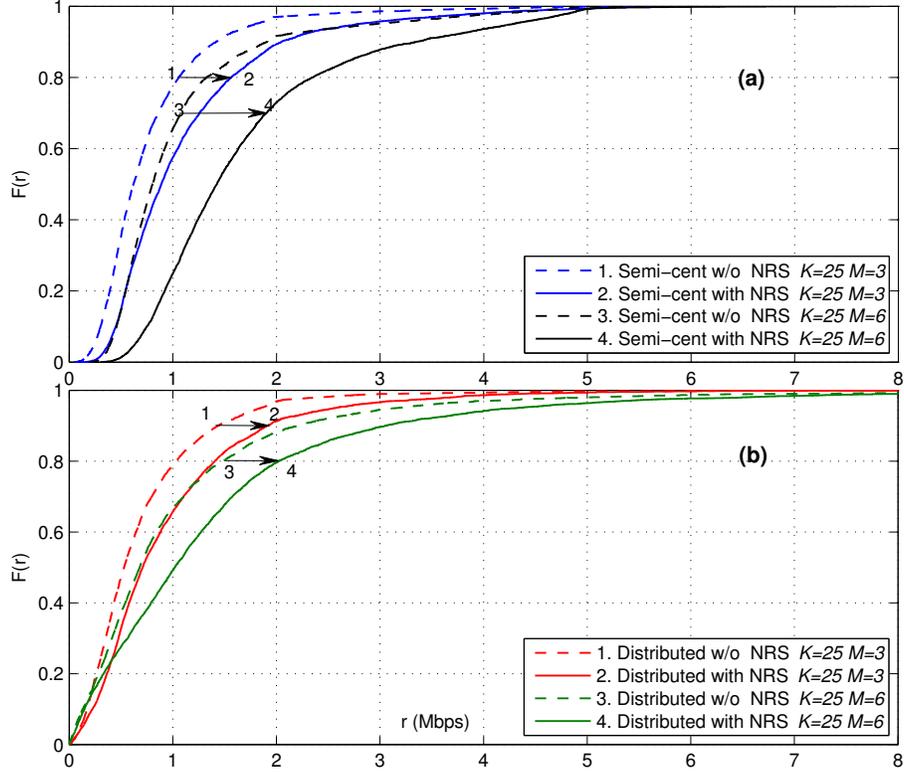


Figure 5.8: CDF of the time-averaged user throughput for the distributed and semi-centralized schemes, using 3 or 6 FRSSs, with and without NRSs, $K = K_{nom} = 25$.

Table 5.2: Outage probability based on users' instantaneous throughput rates, $R_{min} = 250$ Kbps.

Parameters	Semi-centralized		Distributed	
	With NRS	w/o NRS	With NRS	w/o NRS
$K = 15$ $M = 6$	0.0155	0.0280	0.2135	0.2686
$K = 25$ $M = 6$	0.0398	0.1416	0.2669	0.3315

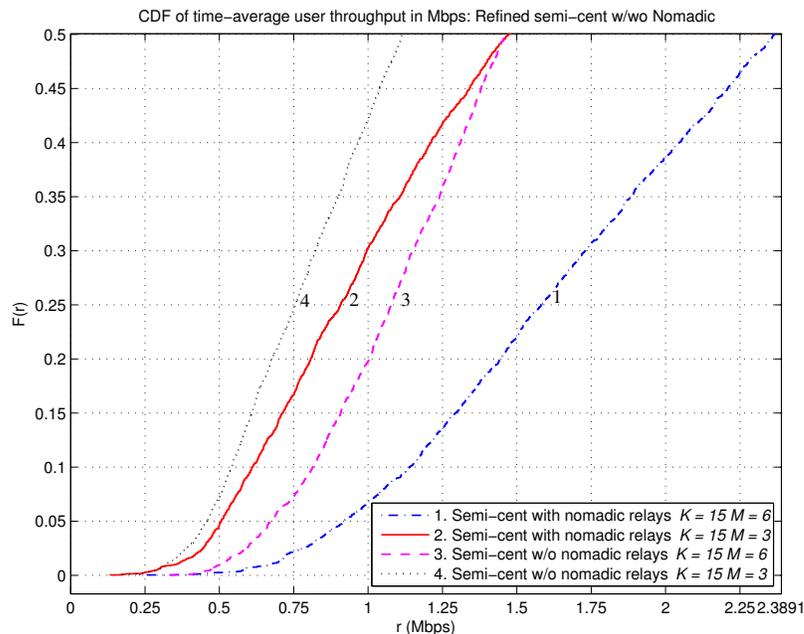


Figure 5.9: Cell-edge performance demonstrated by the lower-tail behavior of the CDFs of the time-averaged user throughput for the semi-centralized scheme with and without NRSs, $K = 15$.

Figure 5.10 shows a scatter plot of user time-averaged throughput as a function of user distance from the BS for the distributed and semi-centralized schemes with and without NRSs, $K = 25$. Each point in the scatter represents the time-averaged throughput for a particular WT within a drop with fixed location and shadowing. It is observed that the scatter plots show a very high throughput for the WTs close to the FRSs. Recall that in the final step of the resource allocation process, the BS assigns the remaining subchannels (after satisfying the minimum required rate of all columns) to the feeder links. Had the remaining resources been assigned to the direct WT, the observation would have been different. The scatter plots give an insight on the coverage improvement throughout the cell area due to the the NRS assistance and attest to the superior performance of the semi-centralized scheme. *The network reuse factor* represents the number of times a subchannel is utilized concurrently (during the same BS sub-frame) all over the cellular network comprising N_c ‘regions’ or cells,

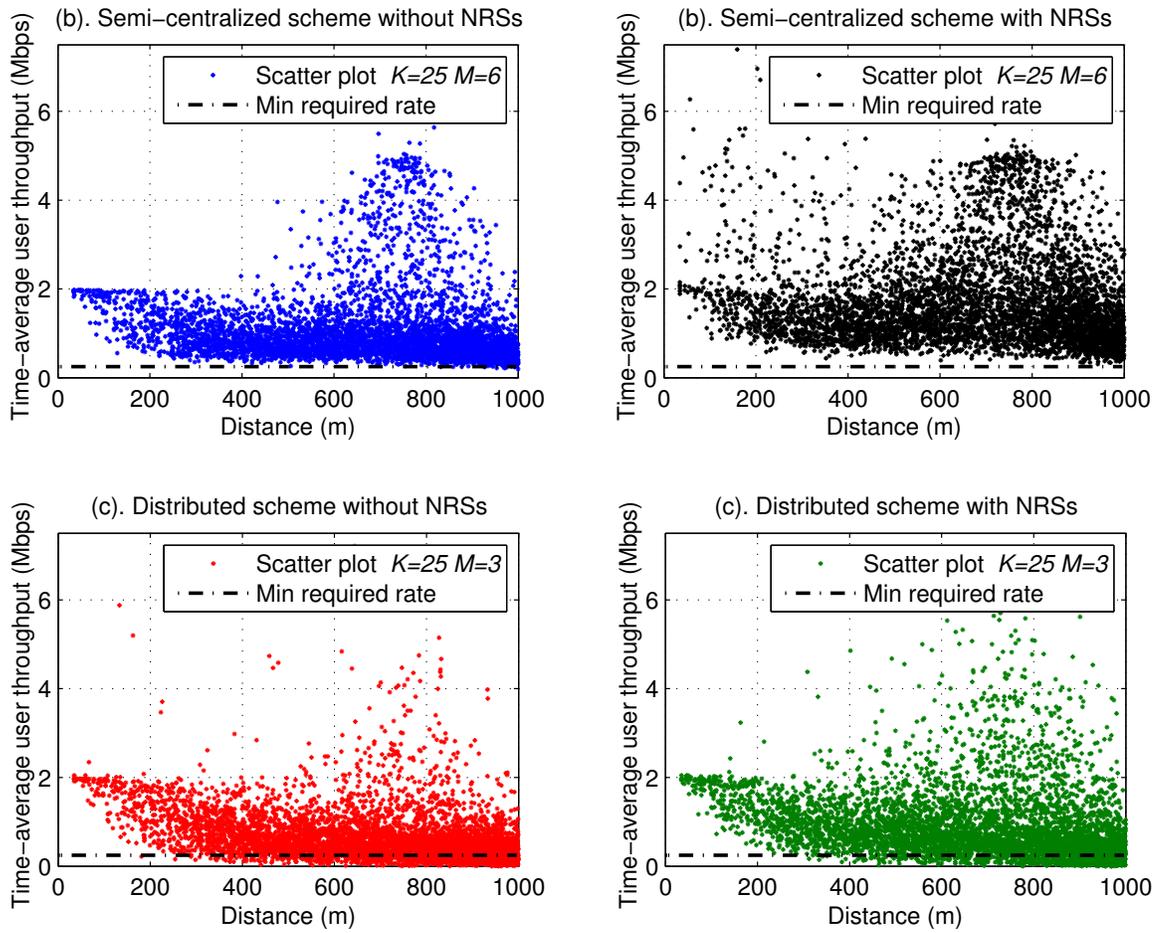


Figure 5.10: Scatter plot of the time-averaged user throughput for the distributed and semi-centralized schemes with and without NRSs, $K = 25$.

and then normalized by N_c . Since all the premium resources are allocated in each cell by the underlying RRA algorithms, the network reuse factor is expressed as follows

$$\text{FR}(n) = \frac{N_c + \sum_c \sum_k \rho_{c,k,n}}{N_c}, \forall n \in \mathcal{N}. \quad (5.6)$$

Where $\rho_{k,n}$ is a binary indicator variable that is set to 1 when NRS $_k$ in cell c acquires subchannel n . Statistics are collected from all subchannels and $N_c = 19$ cells. Note that in the schemes without NRS, a unity network reuse factor should be always realized.

Observing a normalized histogram of such statistics, network reuse realizations of 1.5, 2 and above are obtainable; since the counts are discrete integers, this means that a subchannel is opportunistically reused twice or more in a cell as compared to the static reuse patterns which limit the degrees of freedom in the system. Figures 5.11 and 5.12, for the distributed and the semi-centralized schemes, respectively, present the normalized histograms showing the frequency of channel reuse factor realizations for different configurations of the network parameters, K and M .

Table 5.3 also attests to this fact. Increasing the number of WTs in both schemes, and thus the number of assisting NRSs, results in the occurrence of higher reuse factors. In addition, the dynamic routing strategy in a network with denser FRS deployment results in more relayed users through the FRSs to which NRS cooperation is tied; this means more subchannel reuse. In general, the semi-centralized scheme provides higher average reuse factor than the distributed. Note that the semi-centralized scheme can satisfy the rate requirements using fewer subchannels and therefore more remaining subchannels are assigned to the relayed users who invoke as such the NRS cooperation more often.

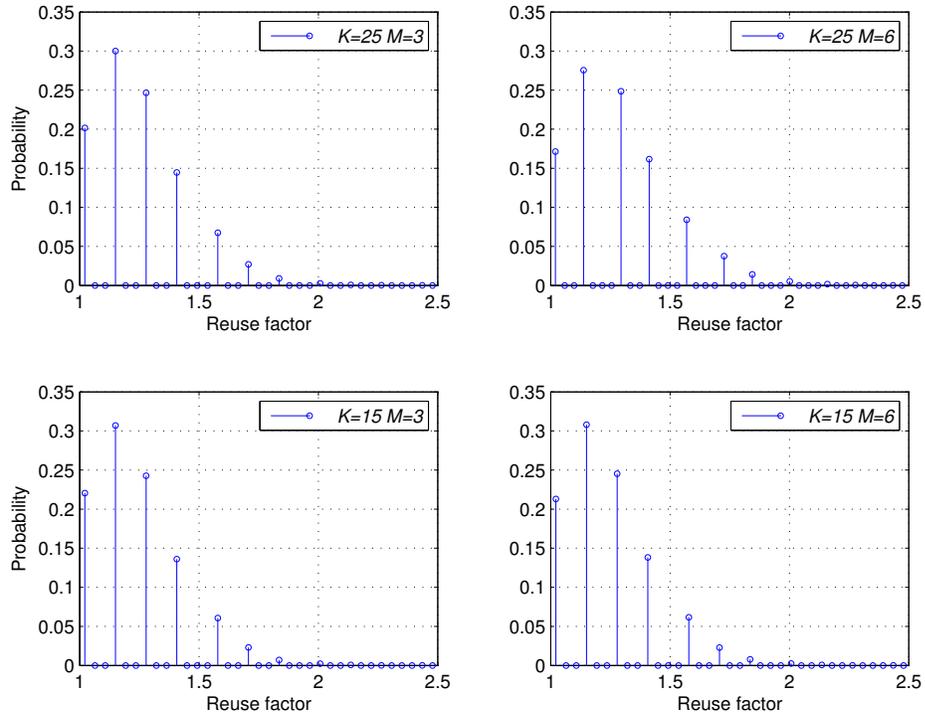


Figure 5.11: Normalized histogram of the subchannel reuse factor for the distributed scheme.

Table 5.3: Mean network reuse factor for the proposed distributed and semi-centralized schemes.

K	M	Distributed	Semi-centralized
15	3	1.2319	1.3156
15	6	1.2351	1.3655
25	3	1.2448	1.3387
25	6	1.2745	1.4282

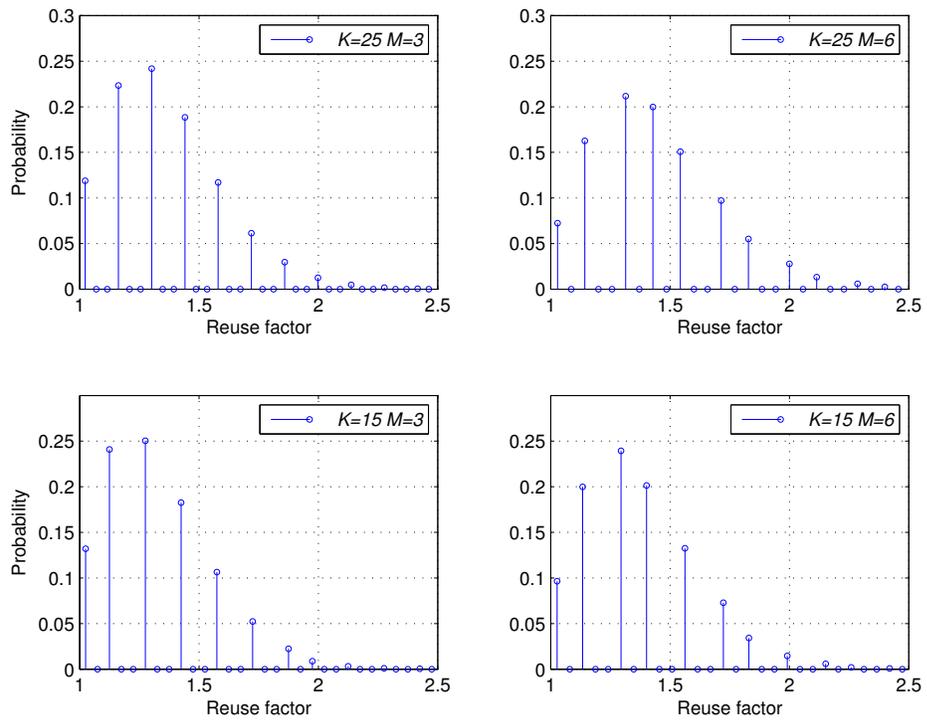


Figure 5.12: Normalized histogram of the subchannel reuse factor for the semi-centralized scheme.

5.7 Conclusions

Next-generation networks, that will comprise a plethora of wireless relay stations of different characteristics will benefit to a great extent from self-organization capability. This chapter describes decentralized RRM methods featuring aggressive and opportunistic reuse in OFDMA-based multicellular networks enhanced with a mix of FRSs and self-organizing NRSs. A novel user-based self-optimizing intra-cell routing strategy that significantly reduces the feedback overhead is employed. We develop novel methods by which NRSs acquire radio resources autonomously without relying on a central entity. Through the asynchronous medium access of the NRSs, aggressive and opportunistic intra-cell resource reuse is attained as opposed to the static reuse patterns often adopted in literature. Furthermore, we introduce a resource-efficient cooperation protocol between an NRS and a serving FRS to assist a troubled wireless terminal. Two underlying heuristic resource allocation schemes of different decentralization levels are devised for the BS and the FRSs. Our extensive numerical results attest to the efficiency of the proposed schemes in terms of time-averaged user throughput, outage probability, as well as network reuse factor. To the extent of our knowledge, no work so far has provided mechanisms for integrating the autonomous NRSs into the cellular network or suggested the underlying RRM schemes and protocols to facilitate their coexistence with FRSs.

Chapter 6

Joint Power and Subchannel Allocation for the Self-Organizing Nomadic Relays in OFDMA-based Cellular Fixed-Relay Networks

6.1 Introduction

Power control (PC) is an important interference combatting mechanism thereby constituting a means for improving the network performance through enabling efficient utilization of system resources. Traditionally, PC has been employed for combatting co-channel and adjacent-channel interferences in multicellular networks. The CCI due to frequency reuse is one of the most limiting factors on wireless system capacity. However, the RRM schemes earmarked for the future wireless networks are designed to be aggressive in frequency reuse.

Centralized PC schemes require reliable measurements of the gains in all radio links in the system. For practical implementation, PC schemes have to rely on far less accurate measurements and limited information to enable distributed operation. A common experience is that, a simple proportional control algorithm, which increases the transmitter power in a link if the received SINR is too low and decreases it when the SINR is high, will normally work well [132]. Thus, such an approach

can be viewed as a gateway to distributed PC, where a transmitter utilizes limited knowledge for power adaptation rather than gathering all network link gains at a central entity. Towards that end, [133] develops distributed iterative PC algorithms that utilize available measurements and converge in stochastic sense, in contrast to deterministic PC schemes that rely on exact or perfect control quantities (SINR or interference). In [134], the authors consider iterative PC as part of the DL schemes employing a non-orthogonal amplify-and-forward cooperative protocol. Each BS independently re-computes user received inter-cell interference as a function of the transmit power values in the previous iteration. Since the iterative process in that work is not guaranteed to converge for all power values, the authors suggest that the distributed algorithm be separated to operate on two independent sets; directly connected and relayed users.

It is known that PC can provide: 1- Means to prevent receiver saturation when the transmitter is too close. 2- Energy savings. Here, the green radio initiatives may readily come to mind [135] while at the user end for instance, low-consumption terminals that depend on solar energy have been already developed and showcased in the Mobile World Congress 2009 [136]. 3- Throughput improvement under frequency reuse through reducing the associated CCI. It will be desirable therefore for a transmitting node, especially if battery-powered, to utilize the minimum energy possible to achieve the desired QoS.

It is worth stating that no matter how smart an RRM scheme is, it will still be essential to formulate strategic energy utilization policy. This is particularly important for dense networks with aggressive frequency reuse schemes such as the one we proposed in Chapter 5 where the battery-powered NRSs are also the entities responsible for the intra-cell reuse. Such a strategic approach may generally represent how the future OFDMA-based relay networks are designed.

While it is known that a relatively marginal gain is attained when employing power adaptation on top of AMC in OFDM networks [137], it is important to note that the NRSs, as adopted in Chapter 5, forward the overheard transmissions while preserving their original AMC modes which are not selected based on the NRS-WT link quality. Therefore, there are potential opportunities in extending our novel schemes to encompass the power adaptation dimension by implementing a joint power and subchannel allocation algorithm at the NRS. The main idea behind that joint algorithm is to allocate to the requested data segment the proper subchannel and the proper amount of power to meet the target SINR of its AMC mode, based on the medium access interference information already available. The objectives of such an add-on is to further mitigate the CCI and provide prudent energy utilization at the battery-powered NRSs without compromising the performance gains before hand. This also brings about the timely environmental concerns and the green wireless initiatives in designing future wireless networks, e.g., [135], [136], and [138].

Interestingly, the adaptive PC is realized in an open-loop manner as no feedback is required by the NRS to function. Thus, we are able to investigate the system to provide answers to questions such as: **How much power savings can be obtained through NRS power control? How much is the throughput gain? Is there any throughput-power savings trade-off?**

The contributions in this chapter can be summarized as follows:

- We present a novel joint power and subchannel allocation algorithm for the self-organizing NRSs in the emerging OFDMA-based fixed-relay networks. The algorithm performs adaptive power control (APC) within the autonomous opportunistic NRS medium access and channel reuse, using two different approaches, to assist a troubled WT through cooperation with the serving FRS.

- The APC mechanism is realized in an open-loop manner requiring no feedback from the assisted WT; significant gains have been achieved through NRS assistance at no additional cost.
- The performance returns in terms of power savings and user throughput are evaluated in a realistic environment considering non-ideal NRS-WT links.
- We identify a throughput-power saving trade-off in terms of the number of deployed FRSs.
- While no other work in the literature has addressed such systems and architecture, we further sustain the concept of nomadic relay-augmented fixed-relay networks we previously established in Chapter 5.

The work presented in this chapter has been presented in the conference paper [123] and it is part of the invention disclosure [124] and the technical report [139]. A journal paper is in preparation.

6.2 System Description

The system model, set of assumptions, as well as the NRS medium access and cooperation functions in this chapter are basically identical to those of Chapter 5.

6.3 NRS Joint Power-Subchannel Allocation Algorithm

The joint algorithm does not lend itself readily to a simple power increase/decrease command as suggested in [133] since the NRS operations encompass selective transmission of multiple AMC modes, medium sensing, and joint power and resource acquisition. Rather, a more elaborate approach is adopted to effectively integrate APC. Recall that the algorithm aims at allocating the proper subchannel and the proper amount of power to the requested data segment such that its target SINR (corresponding to its AMC mode) is achieved, using the medium access interference

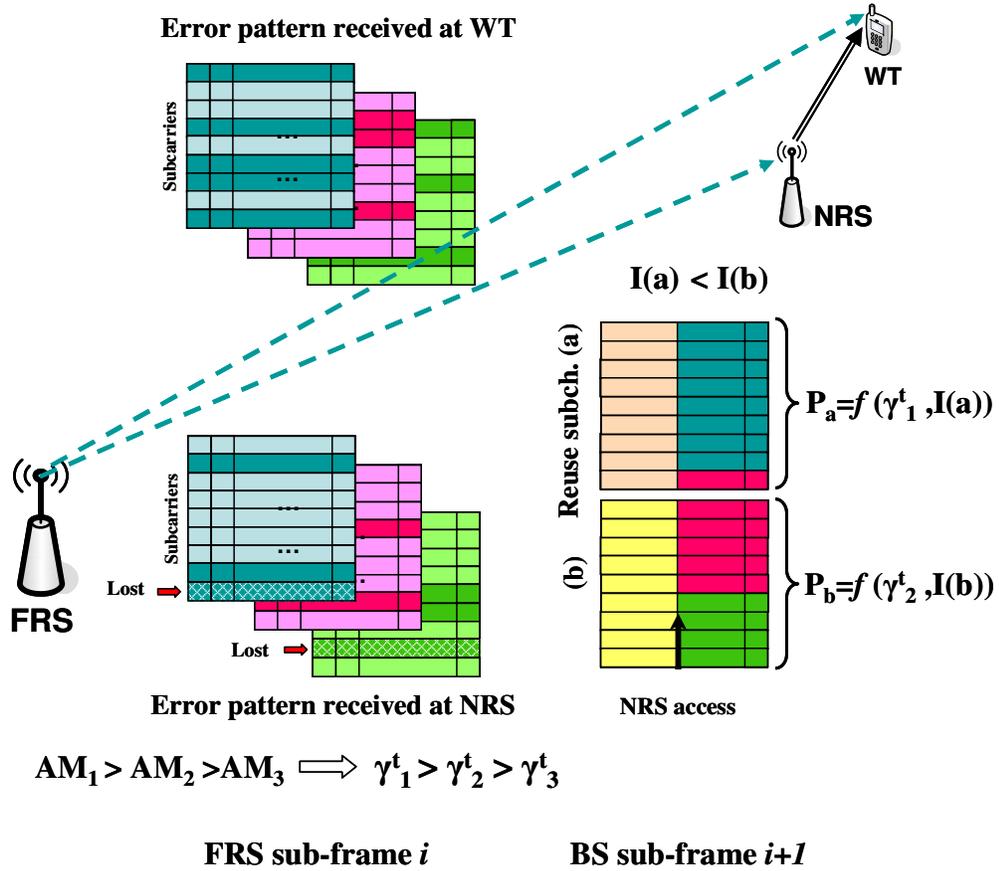


Figure 6.1: NRS power-subchannel allocation for the case where the least possible number of subchannels is used, given the instant of the NRS medium access.

information. Preserving the AMC modes of overheard transmissions simplifies the coordination between the NRS and the assisted WT. Moreover, conventional combining techniques such as selection or MRC can simply be invoked at the WT side, if desired. The two possible approaches to grouping the selected subcarriers are described as ‘subchannel packing’ and ‘no-packing’. Figure 6.1 shows how the subchannel packing is done through example of using three subchannels on the FRS-WT link each possibly of a different AMC mode while the erroneous subcarriers (shown in darker shades) are identified at the WT and the overhearing NRS. During the following BS sub-frame, the NRS forwards the subcarriers it reliably detected out of those requested by the WT. Hence, a subcarrier is lost if it has been requested by the WT while erroneously received by the NRS. Given the shown medium access instant, the NRS acquired two reuse subchannels with the least interference observed. For the

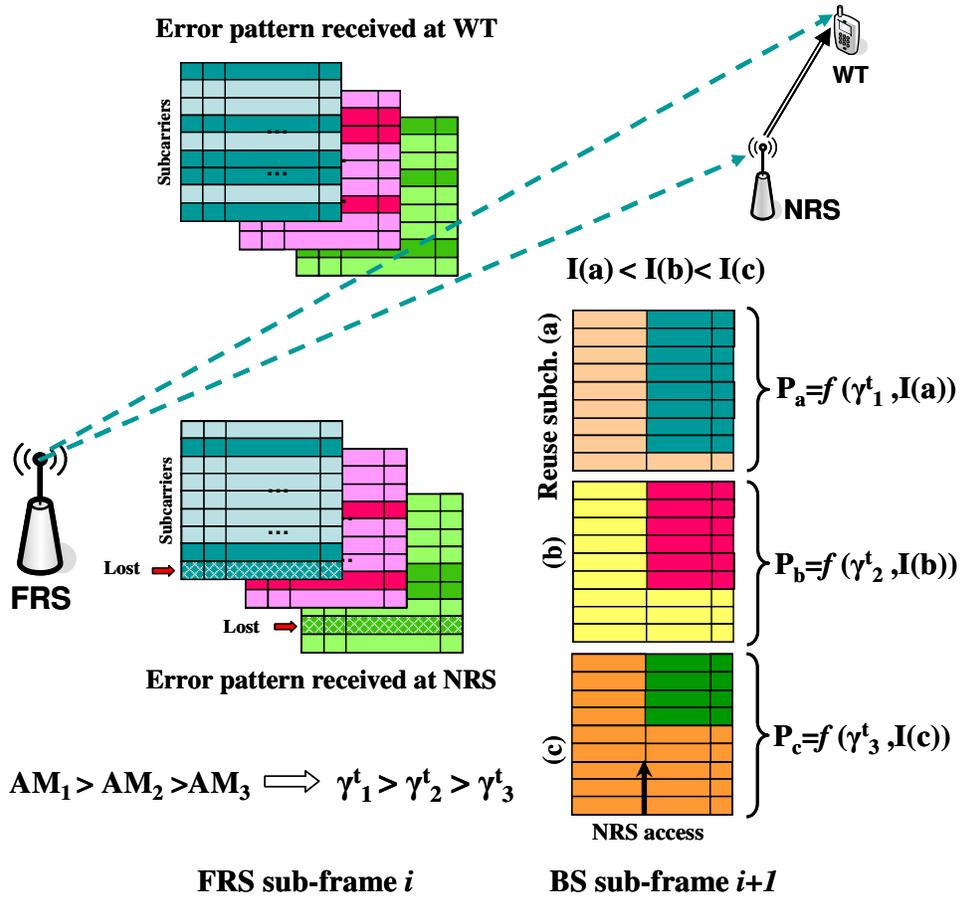


Figure 6.2: NRS power-subchannel allocation for the case where no packing is employed and potentially more subchannels are used.

no-packing mechanism, more reuse subchannels are likely used (e.g., 3 instead of 2 as shown in Figure 6.2). The power calculation, however, is done similarly yet some subcarriers of low target SINR will enjoy a higher SINR if packed with a subcarrier group of high target SINR. While the packing approach has the advantage of utilizing least number of subchannels, it may not necessarily be of less total power consumption. In addition, its physical layer viability would need to be verified. As such, we describe the packing approach for completeness of illustration while it suffices to examine the performance of the no-packing approach to demonstrate the efficiency of the joint power and subchannel allocation algorithm.

In both approaches, NRS_k transmit power on an acquired subchannel \hat{n} is a function of the target SINR of the forwarded subcarrier group i , $\gamma_{k,i}^t$, the interference power

$\mathbf{I}_k(\hat{n})$, as sensed by the NRS during its medium access, and the *large-scale pathloss* P_L which is the same across all the subchannels on that NRS-WT link and can be estimated and/or feedback over quite long time intervals. Saving the instantaneous feedback overhead from the WT reporting the subchannels of the NRS-WT link, the NRS exploits such potentially high link quality and limited variability of the small-scale fading by assuming a certain fading margin F_M . This margin also accounts for the slight interference variability as seen by the WT compared to the sensed $\mathbf{I}_k(\hat{n})$ by the NRS. Let $P_{k,\hat{n}}$ denotes the transmit power of NRS $_k$ (serving WT $_k$) on subchannel \hat{n} while n_0 is the noise power per subchannel. The joint power-subchannel allocation algorithm is the following:

1. The number of (reuse) subchannels to be acquired N_r is determined based on the total number of subcarriers to be forwarded, the remaining portion of the BS sub-frame, and whether or not the packing approach is employed¹.
2. All subchannels other than those assigned to the BS-WT $_k$ link, $\mathcal{N} - \mathcal{N}_{0 \rightarrow k}$, are sorted in ascending order of their observed interference power and the first N_r subchannels are acquired, i.e., $\mathbf{I}_k(\hat{1}) < \dots < \mathbf{I}_k(\hat{n}) < \dots < \mathbf{I}_k(\hat{N}_r)$.
3. The subcarrier groups to be forwarded are sorted in descending order of their target SINRs computed from the spectral efficiencies η_i of the AMC modes as

$$\gamma_{k,i}^t = \frac{(2^{\eta_i} - 1) \ln(5P_e)}{-1.5}. \quad (6.1)$$

4. The first unscheduled group of subcarriers with the highest target $\gamma_{k,i}^t$ are scheduled on the first unassigned subchannel \hat{n} which has the least interference $\mathbf{I}_k(\hat{n})$.
5. If the subchannel packing approach is employed and the subchannel is not fully loaded, more subcarriers from the next quality levels, i.e., $\gamma_{k,i+1}^t, \gamma_{k,i+2}^t, \dots$, are packed into the subchannel (see Fig. 6.1). Otherwise, the subchannel is left

¹For the no-packing approach, $N_r = \sum n_r$ from equations (5.4) and (5.5) for each subcarrier group. For the packing approach, N_r is calculated collectively over all subcarrier groups.

partially loaded (see Fig. 6.2).

6. $P_{k,\hat{n}}$ is calculated in dBm using the following formula;

$$P_{k,\hat{n}} = \min \{P_k^{\text{rem}}, 10 \log_{10}(\gamma_{k,i}^t(\mathbf{I}_k(\hat{n}) + n_0)) + F_M + P_L\}, \quad (6.2)$$

where P_k^{rem} is the remaining amount of the total power available at NRS_k, P_{nom} .

Therefore, P_k^{rem} is reduced thereafter by the calculated amount $P_{k,\hat{n}}$.

7. The steps 4) to 6) are repeated for each of the unassigned subchannels.

Allocating the most SINR-demanding subcarrier group the subchannel with the least interference results in the least amount of transmit power and hence the joint algorithm attempts to achieve the target SINRs and BER in an iterative manner with minimal total power expended and least CCI caused. Note that in the reference ‘NoPC’ NRS-augmented schemes proposed in Chapter 5, the subchannel transmit power P_k^{max} is set to a fixed level that is equal to the total available power divided by the total number of subchannels in the system, i.e., P_{nom}/N .

We note that this novel joint power and subchannel allocation algorithm at the NRS attempts to achieve the target SINRs and BER with minimal total power expended. Figure 6.3 shows a flow chart of the NRS joint power and subchannel allocation algorithm with both approaches integrated.

6.4 System Performance

In this section, we analyze the results obtained through the simulations to demonstrate the efficiency of the proposed joint algorithm for the NRS assistance in the OFDMA-based cellular fixed-relay networks. The distributed RRM scheme presented in Chapter 5 with fixed NRS power per subchannel is considered as our reference scheme in this study and is denoted as the ‘NoPC’ scheme. The PC mechanism is

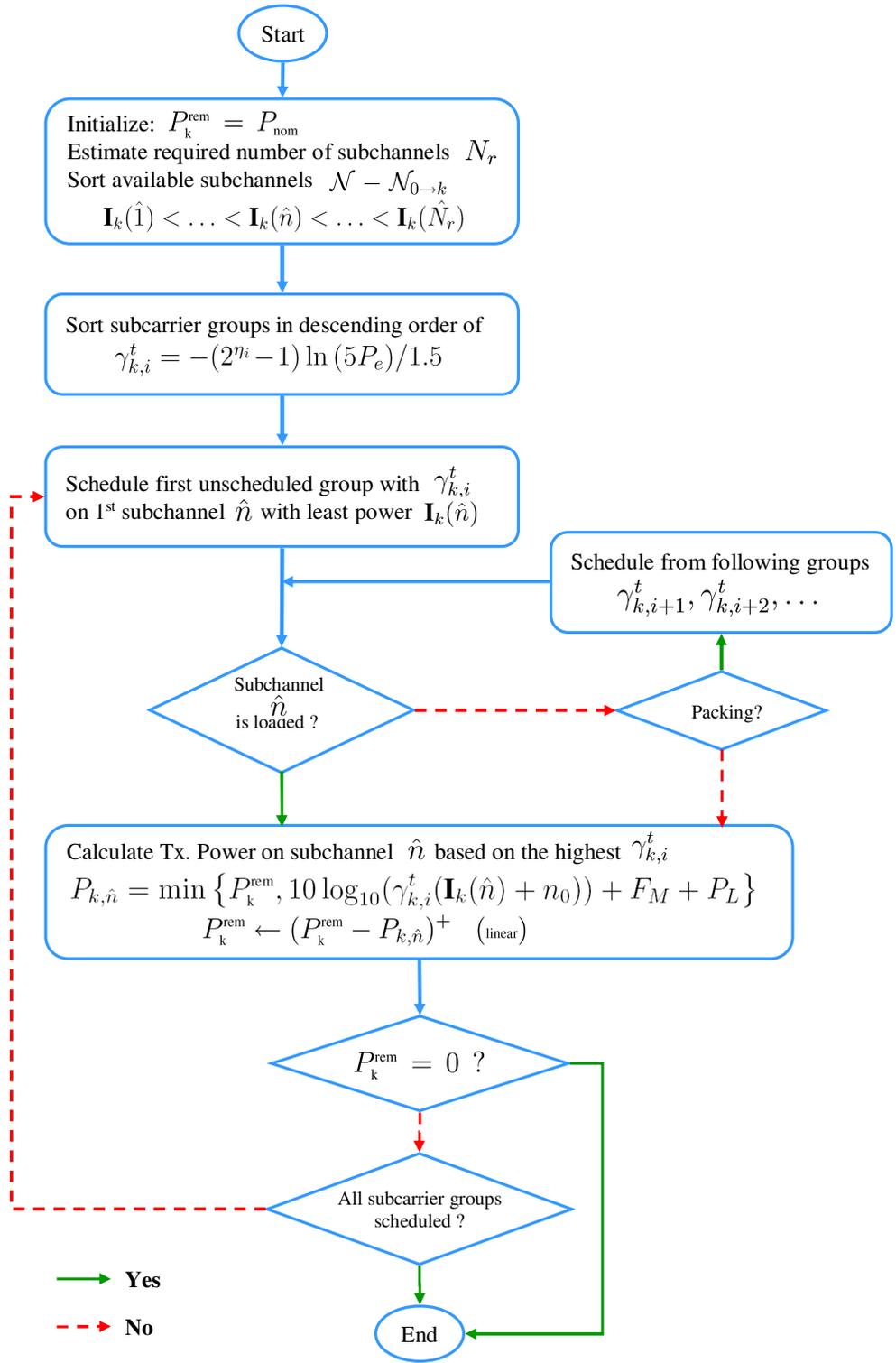


Figure 6.3: Flow chart of the NRS joint power and subchannel allocation algorithm.

expected to affect various aspects of the system performance such as the total power consumption at NRSs, the CCI to the BS transmission during the first sub-frame, and the overall user relayed throughput. Therefore, it is essential to equip the original simulation platform with a module to simulate the links between the NRSs and their associated WTs which were considered in our earlier investigations of the role of NRSs as lossless or error-free. The relevant system parameters are the same as shown in Table 5.1, except for a few parameters which are specified in the following subsection.

6.4.1 Simulation Parameters and Channel Model

The simulated cellular network consists of 19 hexagonal cells enhanced with 3 or 6 FRSs. These relays are placed at a distance of 0.65 of the cell radius from the BS and with a uniform angular spacing. In order to create a more interference limited environment, the network dimensions considered in Table 5.1 are scaled down by two. As such, the distance between two adjacent BSs is 1 Km and the maximum distance between a WT and its assisting NRS is 10 m. The pathloss parameter is set to $A = 20.0$ for serving NRS-WT links. For all other links in the network, $A = 35$. The serving NRS-WT links experience 1.5 dB lognormal shadowing and time-frequency correlated Rician fading with Rician factor of 7 dB. The total transmit powers for BSs, FRSs, and NRSs are 46, 37, and 23 dBm, respectively. Interestingly, these set of parameters have been recently adopted by researchers investigating the performance of femtocells [30]. On that note, our views on extending the NRS medium access and resource acquisition to femtocells are pointed out in Chapter 7.

6.4.2 Numerical Results and Discussions

We start by studying the statistics of the total power consumption at the NRSs for the RRM scheme employing the joint algorithm with power control (the PC scheme)

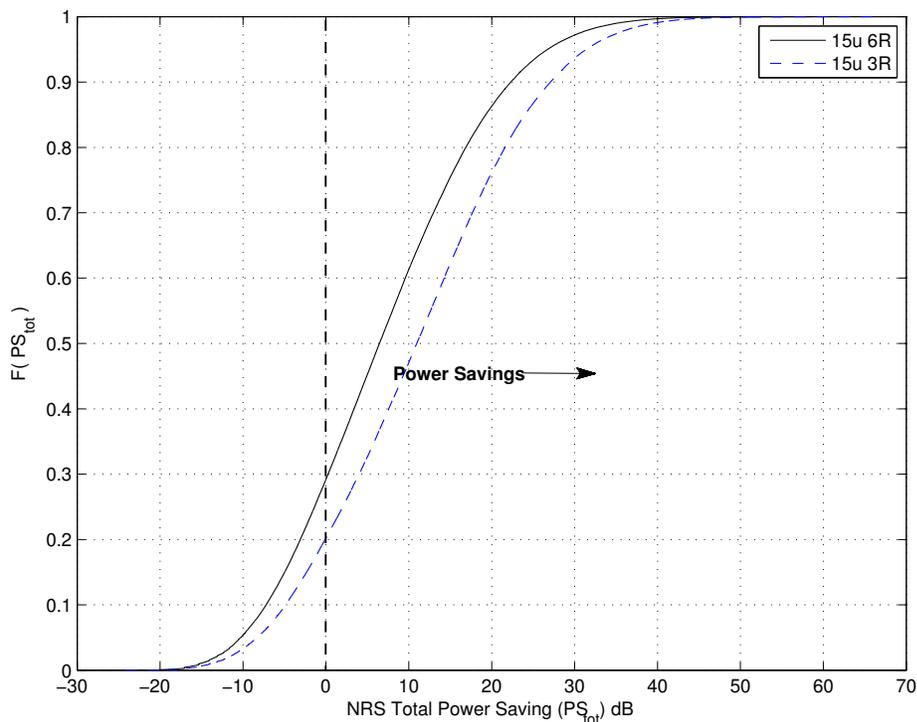


Figure 6.4: CDF plots of the total power saving gain achieved by the RRM scheme with PC for 15 users/cell.

and the reference RRM scheme with fixed power allocation per subchannel (NoPC). We define the saving gain in total transmit power as the ratio in dB of the total transmit power of the NRS with NoPC to that of the NRS with PC.

$$PS_{\text{tot}} = 10 \log_{10} \left(\frac{P_{\text{tot}}^{\text{NoPC}}}{P_{\text{tot}}^{\text{PC}}} \right) \text{ [dB]}. \quad (6.3)$$

Figure 6.4 shows the CDF plots of the power saving gain achieved by the RRM scheme with PC for 15 users per cell while 3 and 6 FRSs are deployed. The wide region to the right of the vertical line at 0 dB represents the saving region where the PC schemes allocate less total power for the NRS transmission than that allocated by their reference NoPC schemes which use the same number of reuse subchannels and a power of 2.91 dBm/subchannel. It can be observed that when 3 FRSs are deployed per cell, the scheme with PC achieves substantial power savings as compared to the NoPC scheme for 80% of the time, while it requires higher total power for 20% of the

time based on the target SINRs, interference, and path loss conditions. The median saving gain in that case is 10.98 dB. Whereas when 6 FRSs are deployed, less saving gains, yet still substantial, are achieved for 71% of the time resulting in an overall median saving gain of 6.44 dB.

An interesting question arises then; why would the PC scheme with 3 FRSs achieve greater total power savings as compared to the PC scheme with 6 FRSs? The answer is not far-fetched; having maintained the same large-scale and small-scale fading statistics for both scenarios, the geographical deployment of the FRSs (with uniform angular spacing) along with the uniform distribution of users over the cell area result more often in larger distances between the WTs and their serving FRSs in the case of 3 FRSs/cell. That implies that FRSs will more often adopt lower AMC modes, which require lower target SINRs, than the case of 6 FRSs/cell. Since the NRS overhears the transmission of the FRS and forwards the selected data using the same AMC modes, the PC expression in (6.2) will result in lower transmit power levels than in the case of 6 FRSs/cell.

This is in fact a sort of power-throughput trade off since our experience with the original NoPC scheme is that deploying more FRSs improves the system performance in terms of user throughput. This still holds for the PC scheme but with less anticipated NRS power savings, as the number of deployed FRSs increases.

It is worth mentioning that another important phenomenon related to the geographical FRS deployment and the novel routing strategy discussed in Chapter 5 is the increase in number of users connected to FRSs as the number of FRSs increases. However, a smaller number of users is expected to be handled by an FRS, yet with higher achievable rates. That implies that the iterative resource allocation algorithm at the BS will satisfy the minimum rates of the feeder links using less resources whereas the remaining resources will be evenly distributed (in terms of number of subchannels) among the feeder links afterwards. As such, the number of subchannels assigned to

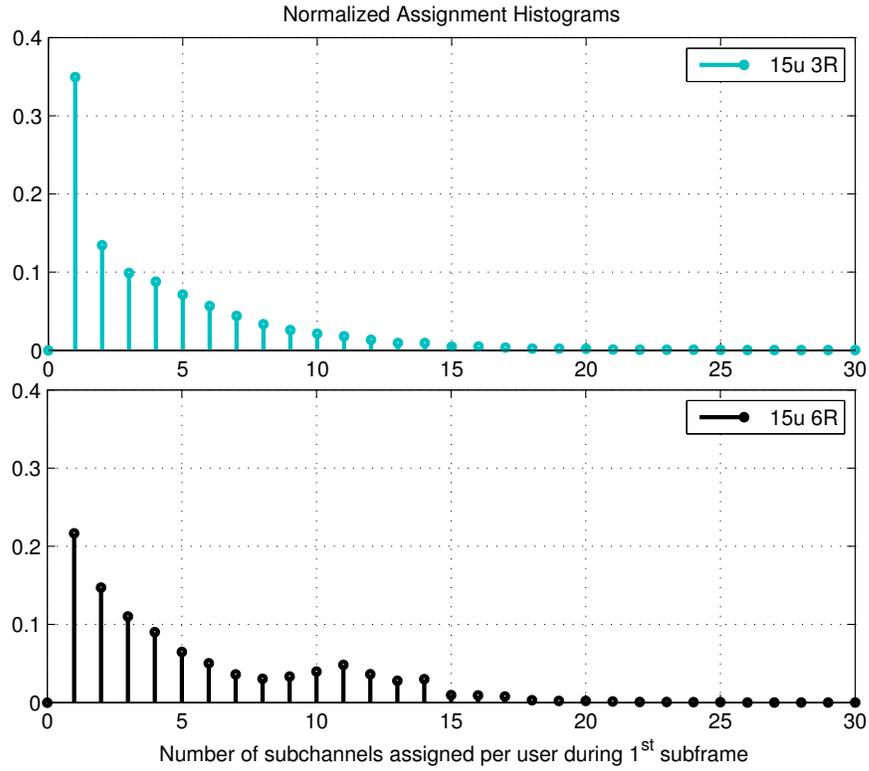


Figure 6.5: Normalized histograms of the number of subchannels assigned to users during the first sub-frame of the DL frame.

an FRS-connected WT can increase as the number of deployed FRSs increases. NRSs are therefore expected to handle more subchannels (from the FRS transmissions) and thus reserve more reuse subchannels. This is evident from Fig. 6.5 which shows the normalized histograms of the number of subchannels (NRS reuse plus BS-WT direct connection) assigned to users during the first sub-frame. It can be observed that the probability mass is further shifted towards the upper tail of the histogram indicating higher subchannel-to-user assignment in the case of 6 FRSs per cell. However, this increase in subchannel assignment applies to both the PC and the NoPC schemes given the same number of FRSs and therefore, it does not influence the statistics of the power saving gain.

In alignment with Fig. 6.4, The CDF plots shown in Fig. 6.6 address the statistics of the power saving gain per reuse subchannel. The definition of the saving gain per subchannel used here is similar to the earlier one in (6.3) as shown in (6.4). However, the CDF plots span larger dynamic ranges as compared to Fig. 6.4. The rare instants

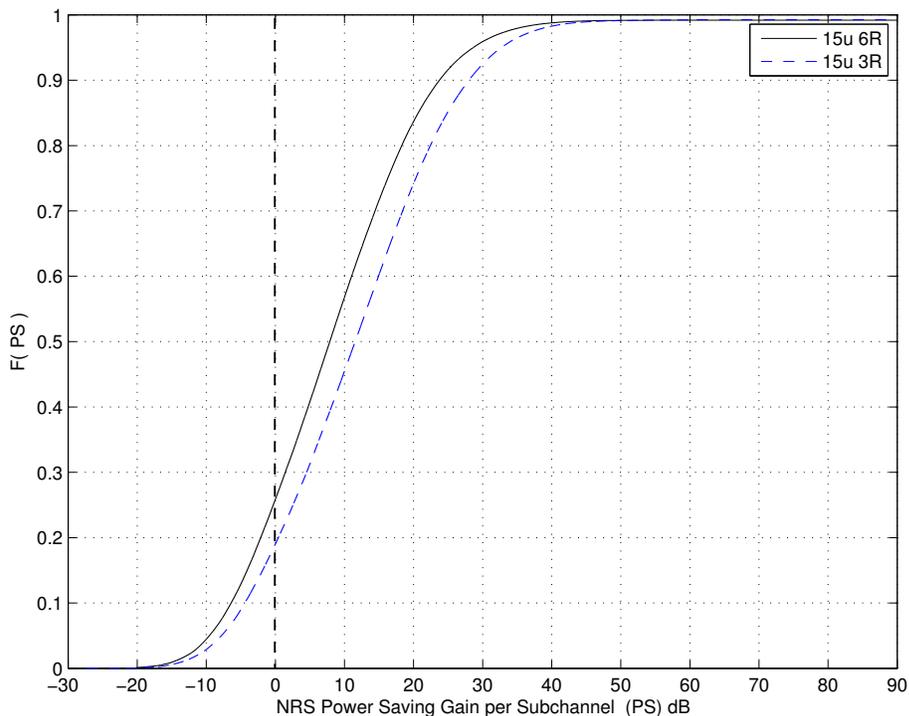


Figure 6.6: CDF plots of the power saving gain per subcahnnel achieved by the RRM scheme with PC for 15 users/cell.

when the PC mechanism allocates very low or zero power to a reuse subchannel are represented by the end of the upper tail of the CDF plots. This is due to the total power constraint P_{nom} which the algorithm satisfies iteratively using (6.2). Although these instants are quite rare, it is interesting to observe that our joint algorithm is designed to satisfy the requirements of the transmissions with the highest AMC modes (target SINRs) first. This means that, when loss is inevitable due to the lack of power, only the last transmissions in the routine, with the lowest AMC modes and hence least number of loaded bits, would be put at risk. This worst-case defense strategy results in the least throughput loss whenever it is inevitable.

$$PS_{k,n} = 10 \log_{10} \left(\frac{P_k^{\max}}{P_{k,n}^{\text{PC}}} \right) \text{ [dB]}. \quad (6.4)$$

We now discuss the impact of power control on user throughput. The NRS transmission occurs during only the 1st sub-frame by reusing the premium system resources

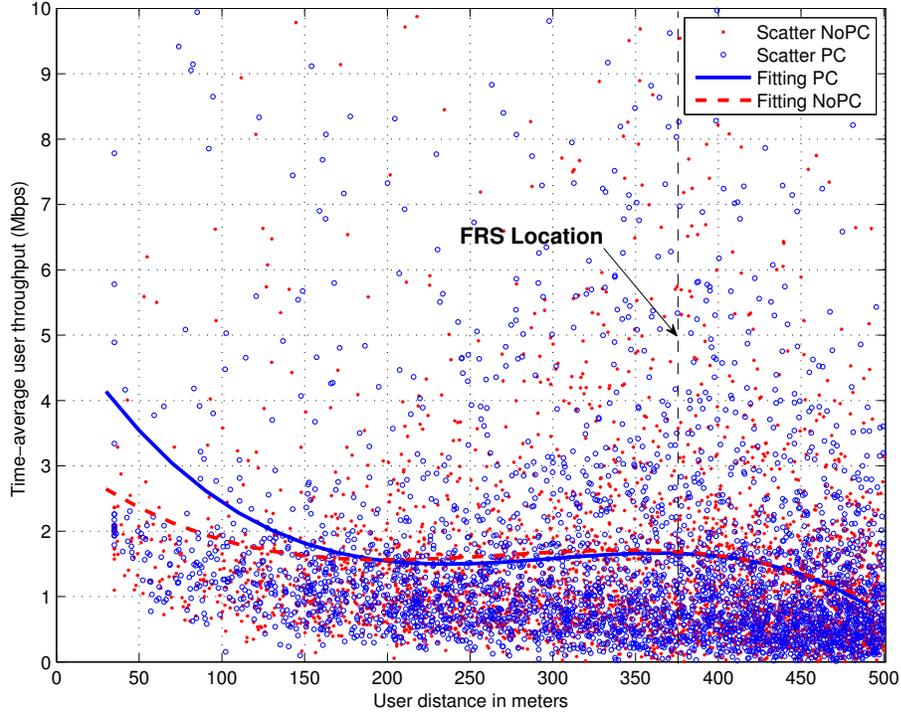


Figure 6.7: Scatter plots of user time-average throughput for the schemes with 15 WTs and 3 FRSs. Curve fittings represent the distance-based conditional means.

used by the BS to communicate with the directly connected WTs and to forward the relayed traffic on the feeder links of the FRSs. The CCI due to the NRS channel reuse may potentially affect the BS-dWT, BS-FRS, or other NRS-WT links. However, FRSs are deployed at strategic locations with good LOS communications and highly directive antennas to the BS; their feeder links are much more immune to CCI than the dWTs. Figure 6.7 shows the scatter plot of user time-averaged throughput versus user distance from the BS for a network with 3 FRSs. As expected, the PC scheme outperforms the reference NoPC scheme within only the vicinity of the BS where dWTs are most likely located and subjected to the CCI due to NRS intra-cell reuse. Observing the vicinity of FRSs suggests that the mutual CCI between NRSs is negligible in the reference NoPC scheme. This is due to the fact that the medium access technique of the NRS, based on listening to the network activity before acquiring resources, has the ability to spread the CCI across the whole bandwidth and reduces the likelihood of reusing the same subchannel among neighboring NRSs. Meanwhile,

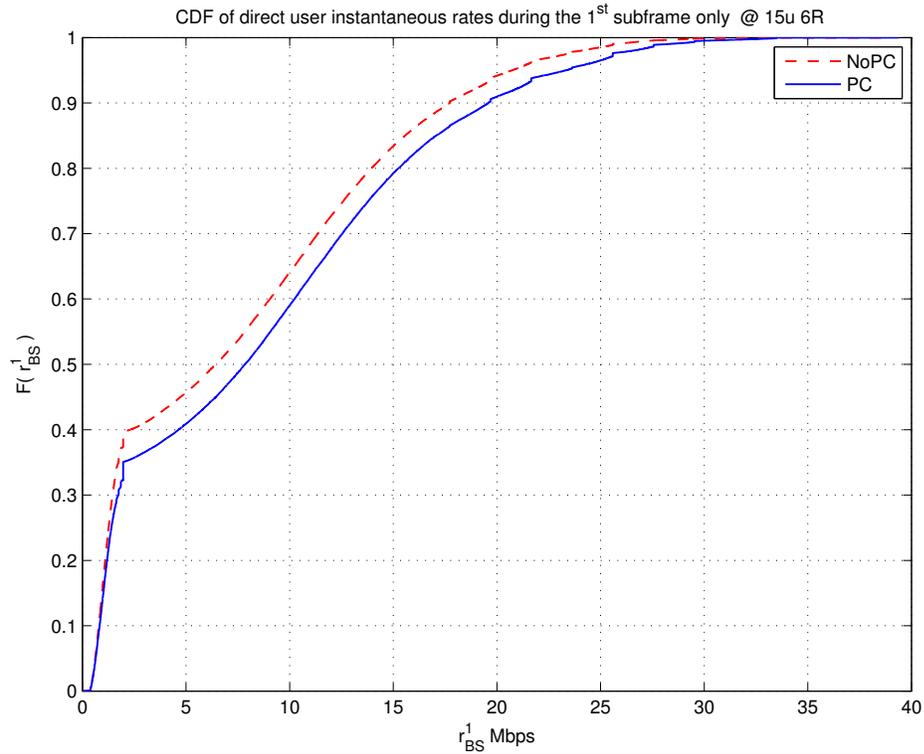


Figure 6.8: CDF plots of the instantaneous throughput of directly connected WTs for the PC and NoPC schemes at 6 FRSs/cell.

the inter-cell CCI, due to NRSs reusing the same subchannel in the adjacent cells, is even less due to the greater path loss given the low transmit power of NRSs.

Since the previous scatter plots and the curve fittings therein comprise points corresponding to both BS and FRS-connected users, it is important to focus on the portion of user throughput subject to the NRS CCI to verify such observations. This is done by performing the statistics on the throughput of only the dWTs throughout the different BS sub-frames and across different drops. For example, figure 6.8 shows the CDF plots of the instantaneous throughput of dWTs for the PC and NoPC schemes at 6 FRSs/cell. It can be seen that for 60% of the time the PC scheme results in a throughput gain of approximately 1.3 Mbps which is in line with our previous observations.

6.5 Conclusions

We devised a novel joint power and subchannel allocation algorithm employing an open-loop APC mechanism for the self-organizing NRSs which are integrated into an OFDMA-based fixed-relay network. The results demonstrate the efficiency of the proposed APC mechanisms. Substantial power savings are achieved at the NRS resulting in extended battery life and less frequent recharging. This smoothly integrated add-on feature to our autonomous NRS medium access technique brings around the timely environmental concerns in designing future wireless communication networks. The value of such an efficient add-on is further emphasized given that the APC mechanism requires no feedback from network nodes and relies on the available medium access information. It has been observed in our illustrative system that APC at the NRSs can further improve the throughput of the BS's directly connected users whereas no throughput gains are achieved for the relayed users due the inherent CCI spreading in the NRS opportunistic medium access. A power-throughput trade-off is addressed in terms of the number of deployed FRSs. We thus further establish our pioneering techniques for realizing the concept of NRS-augmented networks.

Chapter 7

Summary of Conclusions, Contributions, and Possible Extensions

In this chapter we summarize our conclusions and contributions. We also discuss the possible extensions of the work presented in Chapter 3 through Chapter 6 along with some relevant research directions.

7.1 Thesis Conclusions

- Intelligent radio resource management (RRM) schemes are crucial to harness the opportunities in the next-generation relay-enhanced OFDMA-based wireless networks where conventional schemes are not applicable. We address some of the opportunities, challenges, and technical terms associated with the migration from conventional cellular architecture to relay-enhanced. Users' expectations are much higher and fairness obligation is stronger in such networks. We therefore discuss some fairness implementation techniques along with some example fairness metrics towards the design and evaluation of prospective algorithms. Since they are optimal in principle, there are on-going research efforts towards devising more overhead- and complexity-efficient centralized RRM schemes.

Significant savings can be obtained however through distributed schemes despite the potential performance loss.

- In Chapter 3, we have formulated a throughput-optimal policy for an OFDMA-based cellular relay network with symmetric traffic at the BS while the traditional quasi-FDR protocol of that literature is retained. The policy performs joint in-cell routing and scheduling using only two-hop relaying and prevents resource waste, in contrast to prior art, through efficient bit-loading constraints or iterative optimization in the low-complexity algorithm. Comparing the performance under the open routing mode and the practical constrained routing mode, the learning ability of the routing strategy has been demonstrated. However, despite the significant performance returns of that algorithm, it suffers from a performance limiting bottleneck as the traffic load increases. In addition, the quasi-FDR protocol often raises a practical concern due to the limitations in relay hardware technology.
- In accordance with the emerging OFDMA-based cellular relay networks, a novel generalized throughput-optimal formulation employing half-duplex relaying has been presented in Chapter 4. Low-complexity iterative algorithms are devised to solve the separated formulated optimization over two consecutive sub-frames using the queue length coupling. At a slight complexity increase as compared to the quasi-FDR scheme, the network capacity for which the queues can be stabilized has been significantly increased, and hence fairness and ubiquity at high traffic loads, besides the improvement in queue-awareness. Without preset priority weights, service differentiation across classes of asymmetric traffic can be achieved on the time-average and long-term time scales.
- Next-generation networks that will comprise a plethora of wireless relays of different characteristics will benefit to a great extent from self-organization. Under

such high density of relay deployment, decentralized RRM schemes that facilitate the operation and coexistence of the heterogeneous relays and enable aggressive and opportunistic resource reuse have to be considered. Significant savings in feedback overhead can be achieved through a user-based self-optimizing routing strategy given the strategic deployment of the infrastructure fixed relays. Through the self-optimized asynchronous medium access of the NRSs, opportunistic intra-cell resource reuse can be attained as opposed to the traditional channel-unaware static reuse patterns. A Resource-efficient cooperation protocol between an NRS and a serving FRS can be devised based on selective relaying. Our extensive numerical results in Chapter 5 attest to the efficiency of the proposed schemes in terms of time-averaged user throughput, outage probability, as well as network reuse factor.

- Substantial power savings are achieved at the NRS, resulting in extended battery life, through a novel joint power and subchannel allocation algorithm employing an open-loop (without feedback) adaptive power control (APC) mechanism at the self-organizing NRSs. The results in Chapter 6 demonstrate the efficiency of the proposed mechanism. This smoothly integrated add-on feature is in line with the green wireless initiatives in designing future wireless networks. It has been observed that APC at the NRSs can further improve the throughput of the BS's directly connected users in contrast to the relayed users due to the inherent CCI spreading in the NRS opportunistic medium access. On the network-level, a throughput-power saving tradeoff exists in terms of the number of FRSs.

7.2 Thesis Contributions

1. The opportunities and challenges in the next-generation relay networks have been identified and discussed from an RRM perspective in Chapter 2. A comprehensive literature review has been also provided.
2. A study of the performance of throughput-optimal policies in the OFDMA cellular relay network has been conducted in Chapter 3, while retaining the traditional frame structure, through a novel formulation achieving joint routing and scheduling, service ubiquity and load balancing. Unlike existing centralized schemes, substantial savings in complexity and feedback overhead have been achieved.
3. A generalized throughput-optimal formulation has been proposed in Chapter 4 for the practical half-duplex relaying protocol capturing asymmetric traffic, constrained routing and effect of ARQ with and without load balancing. We show that service differentiation is attained through the queue length weights rather than preset weights. Cost assessment of queue-awareness has been provided.
4. Our pioneer decentralized schemes presented and evaluated in Chapter 5 have considered modern architecture comprising high density of relays of different characteristics and functionalities. Novel self-optimized terminal-based routing and NRS medium access have been devised achieving substantial savings in feedback and opportunistic intra-cell reuse, respectively.
5. We present a novel NRS-directed joint power-subchannel allocation algorithm in Chapter 6 performing adaptive power control within the autonomous opportunistic NRS medium access and channel reuse, using two different open-loop approaches (requiring no feedback from the WT). A throughput-power saving tradeoff has been demonstrated in terms of the number of deployed FRSs.

7.3 Extensions to the Centralized Schemes in Chapter 3 and Chapter 4

7.3.1 Intra-cell Resource Reuse

Aggressive resource reuse is a key factor to achieve high spectral efficiencies in next-generation wireless networks. Our proposed RRM schemes in Chapters 3 and 4 use all the available resources in each cell of the cellular network. However, at that stage, intra-cell reuse has not been considered. That is due to the fact that our proposed schemes are intended to address the fundamental resource allocation problem thoroughly seeking, in the first place, efficient management of the system's premium resources. In contrast, the vast majority of schemes in the literature employ intra-cell reuse on top of suboptimal simplifying techniques of allocating the premium resources. Among these techniques are static partitioning of users and relay selection, partitioning of resources among different cell regions or nodes, and excluding the traffic and queue status. As such, the gains from intra-cell reuse therein are potentially consumed to compensate for the incurred spatial, multiuser, frequency, and traffic diversity losses.

Therefore, towards developing more efficient aggressive reuse schemes, we think that extending the proposed schemes to encompass intra-cell resource reuse would be a valuable add-on that can further increase the capacity of these efficient RRM schemes. Our provision to integrate intra-cell reuse can be summarized in the following steps:

1. The algorithm allocates all the subchannels to RS-UT links using the Hungarian algorithm as discussed earlier but without involving the BS. The achievable rates will take into account the interference from BS transmission on all subchannels. Load balancing will be maintained across RSs only.
2. Based on the assignments in Step 1 and taking into account the intra-cell interference, all the subchannels are made accessible to the BS such that subchannels

can be assigned one by one to either a BS-UT link or a BS-RS link, subject to some new constraints. The first set of constraints restricts BS direct transmission to the close UTs in its vicinity. Second, ensures that no BS-RS link is assigned the same subchannel assigned to that particular RS in Step 1 to avoid self interference. The third set of constraints ensures that no UT will be receiving from a RS and directly from the BS on the same subchannel.

Since aggressive resource reuse usually suits the objective of capacity-greedy schemes, it is quite interesting to study the impact of such intra-cell reuse on the performance of our schemes especially in terms of user fairness.

7.3.2 Connection Admission Control for Load Balancing via Inter-cell Routing

We have observed that the proposed fair RRA schemes are capable of distributing available capacity almost evenly among all admitted users regardless of their locations, channels, and interference conditions with minimal impact on total cell throughput. With such behavior realized, and due to the limited resources, it is expected that the fair share of each user will be reduced as the number of admitted users increases. As such, the RRM schemes have to work in conjunction with a connection admission control (CAC) mechanism that decides, based on connected users' QoS, when to admit more users to a particular cell (BS) and when to deny an incoming connection and handover the user to a neighboring cell through a handover mechanism. With the deployment of relays, more handover opportunities arise through enabling inter-cell routing. We believe that such CAC and handover mechanisms would be essential and integral parts of the prospective RRM schemes towards practical implementation.

7.3.3 Partial Feedback: How Much and How Often is Enough?

The results in Section 4.6 show almost no performance degradation in both user throughput and fairness when 50% less feedback is provided from each UT instead of the full feedback (reporting all subchannels). That is due to the fact that our dynamic routing strategy, either in the open or the constrained mode, allows the UT to be connected to more than one node simultaneously; having many users per cell, this implies that only very few subchannels are used per node-UT link. As such, with potentially marginal performance losses, further savings in overhead can be achieved if UTs report only the ‘best’ fraction of subchannels in term of achievable rates. Whether reporting a fixed number of the best subchannels or every subchannel whose quality is above a certain threshold, such number or threshold should be determined as a function of the number of UTs and RSs.

We have also explained that the feedback can be acquired as frequently as each $\lfloor T_c/T_F \rfloor$ frames. That is the maximum integer number of TDD frames less than the user’s coherence time of the channel (4 TDD frames in the simulated scenario). Taking the CCI in the multicellular network into account, the implication is that the RRA algorithm can be invoked that often while the allocation result will be applied to the transmissions of the intermediate frames until the following allocation instant. Such relaxed resource allocation, however, less exploits the traffic diversity for highly burst traffic. Therefore, it is interesting to relate the minimum feedback frequency to user mobility and the burst nature of traffic.

7.4 Extensions to the Decentralized RRM in Chapter 5 and Chapter 6

7.4.1 The WT-based Self-optimized Dynamic Routing: What Are the Proper CSI Averaging Percentile and Time Window?

In our distributed and semi-centralized RRM schemes, the WT dynamically selects its serving node based the link's quality measure. Such a quality measure represents the best subchannels on that link as the average of the 10th percentile of all the potential instantaneous achievable rates. Although intuitive, such an approach is so far empirical since it is not yet clear how sensitive the performance would be to different averaging percentiles and time windows (greater than one frame). We observe that studying the impact of these parameters is of great value to the literature of distributed RRM in multicarrier environments. The number of WTs, FRSs, and user mobility are perceived the most relevant factors influencing the tuning of these two parameters.

7.4.2 Characterization of the FRS Feeder Link Support to the Allocated Rate on the WT's Access link for Self-optimized Routing

Given the setup in Chapter 5 which exploits the special characteristics of FRSs, we have relied on the assumption that the feeder link of FRS_{*m*} would be of sufficient quality so that the sum-rate on the user's access link is the two-hop end-to-end capacity, i.e.,

$$R_{0 \rightarrow m, k} = \sum_{j \in \mathcal{N}_{0 \rightarrow m, k}} R_{0, m, j} \geq \sum_{i \in \mathcal{N}_{m \rightarrow k}} R_{m, k, i} = R_{m \rightarrow k}, \quad \forall m \neq 0.$$

It is particularly interesting to understand the impact of channel statistics on the likelihood of violating this condition when it comes to facilitating and designing distributed and self-optimized routing and resource allocation in multihop networks in

general. Assuming any non-decreasing achievable rate function of the SNR, the analysis should lead to expressions for $\mathcal{P}\{R_{0 \rightarrow m,k} < R_{m \rightarrow k}\}$ in two cases: 1- Single user connected to FRS_m (worst case). 2- Multiple users connected to FRS_m as shown in Fig. 5.3.

Since it has been cumbersome to analyze the SNRs capturing the statistics of both the small-scale and large-scale fading, using the composite fading generalized-K models developed in [140] would lead to a more tractable analysis. The generalized-K model results from using the Nakagami PDF to model multipath fading and the Gamma PDF to model shadowing. Interestingly, the approach introduced in [140] can be used to well-approximate the distribution of the sum of independent generalized-K random variables by a Gamma distribution.

7.4.3 Performance of the PC Mechanism with Subcarrier Packing

The subcarrier packing approach is described in Fig. 6.1 where a potentially smaller number of reuse subchannels is required as compared to the no-packing approach considered in our study. The benefit of such approach should be realized in the reduction of the CCI during the 1st sub-frame. However, the total power consumption of the packing approach has to be investigated through different simulation scenarios. That is due to the fact that the optimized subchannel transmit power is the same for all the comprised subcarriers and it is calculated based on the most rate-demanding subcarriers.

7.5 Development of the NRS MAC and PC Techniques in Chapter 5 and Chapter 6 towards femtocells

The femtocell concept is steadily attracting research interest in both academia and industry and thus takes part in the envisioned next-generation wireless networks. The

main idea is to bring the high-data-rate capacity to numerous femto-BSs deployed within user hot spots via the wired infrastructure, i.e., DSL, cable, or fiber-optic connections [30]. Such modern radio access network (RAN) architecture is expected to offer reliable broadband connectivity with QoS guarantees that can not be foreseen, for instance, through WiFi as a contention-based RAN operating in the unlicensed band. The immediate implication is that intelligent and distributed RRM is required to smoothly integrate the femto-BSs into the service provider's wireless network. We observe that the medium access technique and the power control mechanisms devised for NRS autonomous operation can be further developed to facilitate the transparent operation of the femto-BS. Our initial thought is that a femto-BS would be serving multiple WTs and therefore the MAC technique and the PC mechanism beforehand have to be consolidated with a feedback-aided scheduler.

7.6 Terminal-to-terminal Relaying and Ad-hoc Networks

An interesting aspect of the envisioned next-generation wireless networks is how to harness the inherent opportunities in the multiuser environment through terminal-to-terminal cooperative relaying. Despite the existence of the infrastructure in the service provider's cellular network, enabling such cooperative relaying protocols imposes the ad-hoc nature on the wireless network.

Most of the works in the literature of cooperation consider non infrastructure-based networks and focus on improving the end-to-end link quality in isolation from the resource allocation. Whereas the literature of resource allocation have not yet adequately addressed cooperative transmissions [141]. Therefore, we observe research merits in bridging the efforts of these two research campaigns to fully exploit the cooperative transmission paradigm. It is also worth mentioning that the work presented in Chapters 5 and 6, would be among the leading works addressing cooperative

relaying within the framework of resource allocation and perhaps the first to study that in the context of nomadic relay assistance.

Currently, we have not yet developed initial thoughts or conducted sufficient literature review to tackle such a research problem. Should we gear our efforts towards that direction, several key issues will constitute our research basis. These issues are summarized in [141] in the form of two questions as follows: 1) Relaying terminal selection or “Who should help whom” among the randomly located users? 2) How many and which resource units (i.e., power and subcarriers) should be utilized for cooperation? The optimization objective in such ad-hoc-like network will aim at minimizing the system power consumption rather than throughput since the terminal battery life is a major concern [137].

Appendix A

Generating a Large Number of Independent Time-Frequency Correlated Fading Realizations in OFDMA Multicellular Networks

As discussed earlier in Section 2.9, radio propagation models have significant impacts on the performance of algorithms designed for wireless communication systems. To test new algorithms, the channel models need to be versatile to adequately represent the real-life environments in which the systems will be operating. Since correlation exists in practical small-scale fading models where the channel gains on a certain link are time-invariant within the coherence time T_c and frequency-invariant within the coherence bandwidth B_c , time-frequency correlated Rayleigh and Rician channel models are adopted in this thesis. However, in system-level simulations, techniques based on FFT/IFFT, such as the Jakes' model, are usually employed to generate an **offline** N -by- N_s channel gain matrix **for each potentially serving or interfering link in the network** similar to that shown in Fig. A.1 (N_s is the simulation drop length in time samples or frames).

Looking at the network layout in Fig. A.2 demonstrating 19 cells with a relatively low user density ($K = 10$), the number of potential serving links for each user is 7 whereas the number of potential interfering links can be calculated by connecting

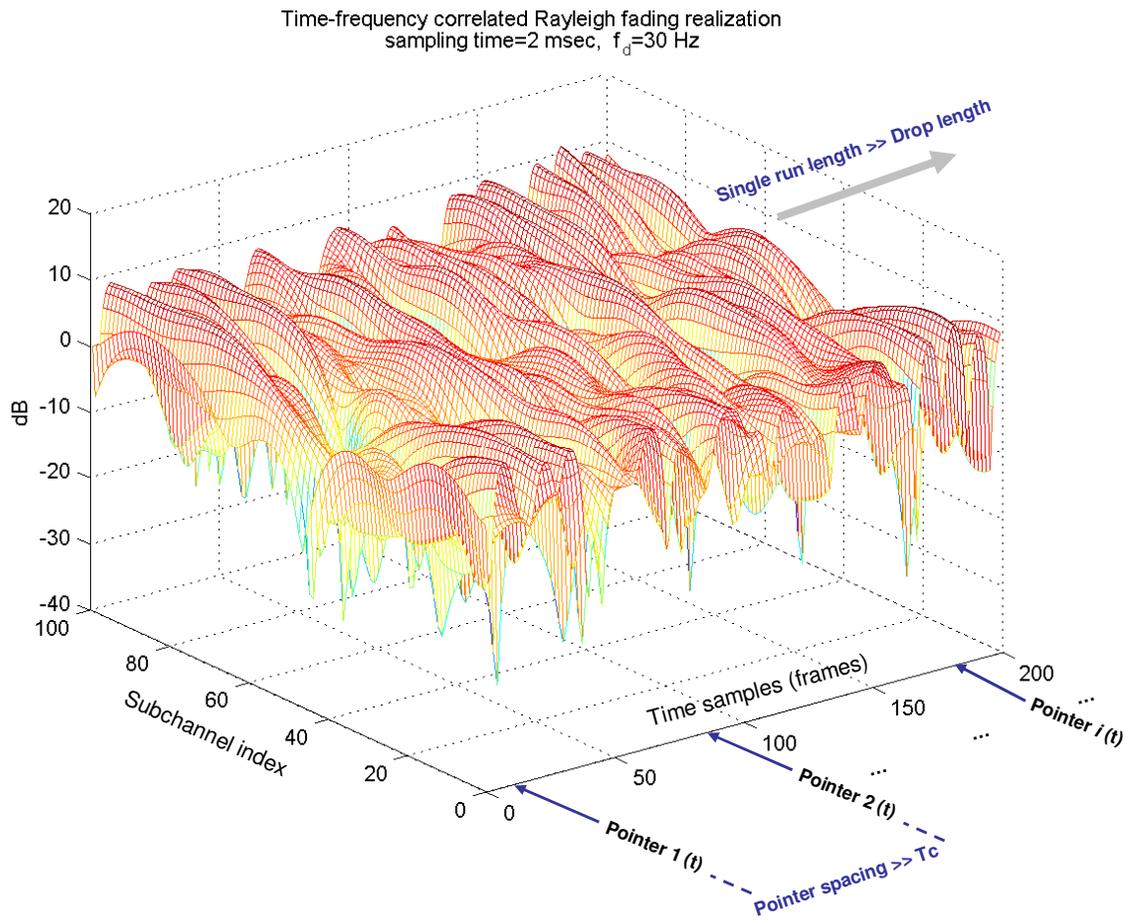


Figure A.1: A single run of the time-frequency correlated fading in a simulation drop showing the pointers' positions at time frame t .

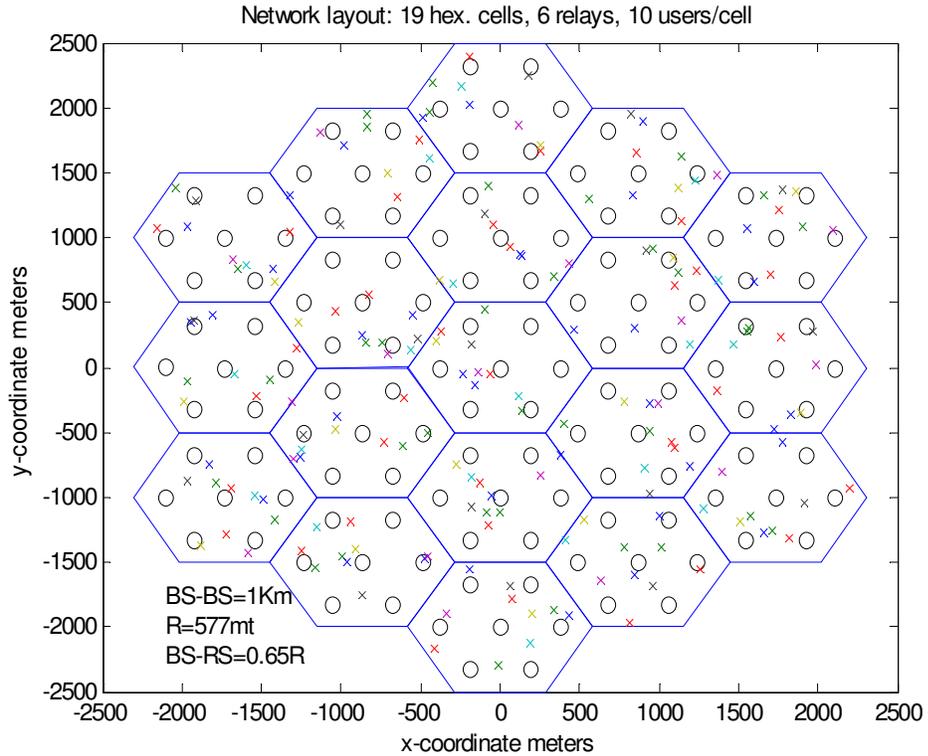


Figure A.2: An example network layout in a simulation drop.

each ‘x’ in the figure to every node in the first and second tiers of cells. That is $18 \times 7 + 7 = 133$ independent N -by- N_s fading matrices for each user in a wrap-around simulation. The total number of 2-D realizations we need in the wrap-around case (preferably for each drop) is generally $(19)^2 K(M + 1)$ which is up to 101080 2-D realizations for the high density scenario of $K = 40$ and $M = 6$. A similar calculation can be done for each FRS in the network but with only one potential serving link. To make it worse, these numbers will further explode once we consider the additional links associated with K_{nom} NRSs per cell. As such, generating and storing that many independent realizations given all the required random samples and FFT computations for each, especially for a large number of subchannels and very long drops, leads to unnecessarily intense computational complexity and wasteful RAM consumption which even might preclude the actual simulations. We also note that loading each realization from a hard drive whenever a sample is needed (to save the RAM) will significantly slow down the simulations which already take several

days to conclude without the channel gains' generation time.

Therefore, the following efficient technique has been used throughout this thesis:

1. For each category of links (specified by a certain power delay profile, Doppler spread, and fading PDF), only one N -by- N_{run} realization is generated at the beginning of each drop, where $N_{run} \gg N_s$.
2. Each serving or interfering link in the network that is intended to fall in one category is represented by a pointer on the time axis of the corresponding realization.
3. The spacing between these pointers is set to be much greater than the coherence time of the channel, i.e., $\Delta t \gg T_c$, where $T_c \approx 0.423/f_d$. As such, at any time sample or frame t , the retrieved channel gain vectors by the individual pointers are uncorrelated.
4. At the time sample $t + 1$, all link pointers are incremented by one sample to maintain the time correlation on their respective links. If a pointer reaches the end of the run length, it continues to retrieve samples from the beginning in a wrap-around manner using the mod function.

Note that:

- Mapping of pointers to respective network links can be easily done in some interleaved manner so that the adjacent pointers i and $i + 1$ correspond to two links that are spatially apart in the multicellular network to further avoid any implications of spatial correlation between the co-located links.
- Due to the wrap-around retrieval of time samples, a violation of the time-correlation between two consecutive samples might occur once per drop on some links. However, the impact of that is negligible given the long drop length.

Appendix B

Channel Emulation and Generation of Error Patterns at the WTs and NRSs with lossless NRS-WT links

We have discussed in Chapter 5 how the NRS cooperate with FRS transmission to a WT through selective relaying which means that the NRS may assist only with the data segments (subcarriers) that it received correctly and while the WT did not (see the reproduced Fig. B.1). In order to implement such a channel effect into our simulation platform to allow for the operation of the LARA protocol, the channel emulation Matlab code shown in Fig. B.2 has been used to generate the random error pattern for each subchannel transmitted by the source (FRS or BS) as seen by the WT and the assisting NRS, as applicable.

The main idea is that whenever an AMC mode with certain spectral efficiency η is chosen based on the feedback SINR γ targeting the BER P_e , a different SINR could be observed at the receiving node(s); such as the NRS which has a link budget that is different from that of its WT. Therefore, the effect of the channel on the BER can be emulated by evaluating the new BER \acute{P}_e based on the new SINR $\acute{\gamma}$ using the same AMC equation (5.1). Assuming independent bit errors, a subcarrier is received erroneously if at least one bit is in error which is a pessimistic approach assuming

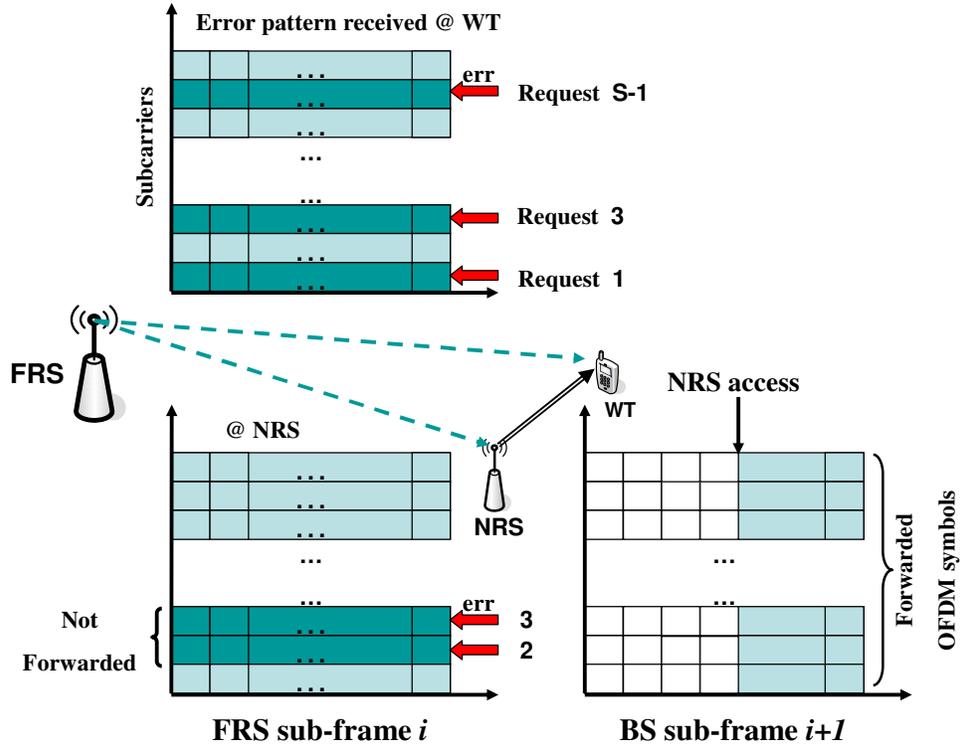


Figure B.1: Reproduced: Illustration of the cooperation of the serving FRS and the dedicated NRS to assist the WT. S is the number of OFDM subcarriers per subchannel. Red arrows indicate the subcarriers received in error at the WT and the NRS.

no forward error correction (FEC) capability for coded modulations. As such, the probability of subcarrier error can be calculated using a version of the packet error probability formulae in [7] as follows

$$P_{sub} = 1 - (1 - \hat{P}_e)^{n_{sub}}.$$

A Bernoulli sequence can thus be generated using P_{sub} to represent the error pattern. A more sophisticated code is considered however in Chapter 6 in which we take into account the effect of a non-ideal NRS-WT link.

```

function [NymbFwd,b]=ErrorPatternGen(tx,q,sinr,sp,W,Tchunk,sinrNom)
% tx=1 --> BS is the source, an FRS otherwise
% q Buffered data at the source
% (new) sinr --> new BER (Pe) different from the target one
% sp chosen spectral efficiency at the source node (AMC mode)
% W subchannel bandwidth
% Tchunk = 24 OFDM symbols --> 12 OFDM symbols per sub-frame*18
subcarriers
% sinrNom received sinr at the overhearing NRS
%
% NymbFwd number of OFDM symbols to be forwarded by NRS
% b number of bits contributing the net throughput
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
Pe=0.2*exp(1.5*sinr/(1-2^sp));
bpsubc=floor(sp*W*Tchunk/(2*18)); % number of bits per subcarrier
Nsubc=min(floor(q/bpsubc),18)*(tx~=1)+18*(tx==1);
Psubc=1-(1-Pe)^bpsubc;
u=rand(1,Nsubc);
errpattern=(u<=Psubc);
numerr=sum(errpattern);
b=bpsubc*(Nsubc-numerr);
NymbFwd=0;
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% NRS assists with the FRS transmissions only in the %%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% considered scenario %%%%%%%%%%%%%%
if (tx~=1)&&(numerr>0)
    PeN=0.2*exp(1.5*sinrNom/(1-2^sp));
    PsubcN=1-(1-PeN)^bpsubc;
    u=rand(1,Nsubc);
    errpatternN=(u<=PsubcN);
    NymbFwd=12*(numerr-sum(errpattern.*errpatternN));
    b=b+bpsubc*NymbFwd/12; % NRS assists and contributes to the net
    % throughput by the number of symbols it can correct
end

```

Figure B.2: Matlab code for error pattern generation (channel emulation) and throughput calculation at the receiving WT and NRS assuming lossless NRS-WT links.

Bibliography

- [1] H. Yanikomeroglu, “Fixed and mobile relaying technologies for cellular networks,” *Second Workshop on Applications and Services in Wireless Networks (ASWN’02)*, pp. 75-81, July 2002.
- [2] R. Pabst, B. Walke, D. Schultz, P. Herhold, H. Yanikomeroglu, S. Mukherjee, H. Viswanathan, M. Lott, W. Zirwas, M. Dohler, H. Aghvami, D. Falconer, and G. Fettweis, “Relay-based deployment concepts for wireless and mobile broadband cellular radio,” *IEEE Communications Magazine*, 42(9), pp. 80-89, September 2004.
- [3] G. Li and H. Liu, “Downlink dynamic resource allocation for multi-cell OFDMA systems,” *IEEE Vehicular Technology Conference (VTC-Fall)*, vol. 3, pp. 1698-1702, October 2003.
- [4] G. Li and H. Liu, *OFDM-based Broadband Wireless Networks: Design and Optimization*, Wiley, October 2005.
- [5] J. Jang and K. B. Lee, “Transmit power adaptation for multiuser OFDM systems,” *IEEE Journal on Selected Areas in Communication*, vol. 21, pp. 171-178, February 2003.
- [6] S. Pietrzyk and G. M. Janssen, “Radio resource allocation for cellular networks based on OFDMA with QoS guarantees,” *IEEE Global Telecommunications Conference (Globecom)*, November 2004.
- [7] D. Niyato and E. Hossain, “Adaptive fair subcarrier/rate allocation in multirate OFDMA networks: Radio link level queuing performance analysis,” *IEEE Transactions on Vehicular Technology*, 55(6), pp. 1897-1907, November 2006.
- [8] W. Nam, W. Chang, S.-Y. Chung, and Y. Lee, “Transmit optimization for relay-based cellular OFDMA systems,” *IEEE International Conference on Communications (ICC)*, pp. 5714-5719, June 2007.
- [9] G. Li and H. Liu, “Resource allocation for OFDMA relay networks with fairness constraints,” *IEEE Journal on Selected Areas in Communications*, 24(11), pp. 2061 - 2069, November 2006.
- [10] **M. Salem**, A. Adinoyi, M. Rahman, H. Yanikomeroglu, D. Falconer, Y.-D. Kim, E. Kim, and Y.-C Cheong, “An overview of radio resource management in relay-enhanced OFDMA-based networks,” *IEEE Communications Surveys and Tutorials*, 12(3), pp.

422-438, Third Quarter 2010.

- [11] **M. Salem**, A. Adinoyi, H. Yanikomeroglu, and D. Falconer, "Opportunities and challenges in OFDMA-based cellular relay networks: A radio resource management perspective", *IEEE Transactions on Vehicular Technology*, 59(5), pp. 2496-2510, January 2010.
- [12] **M. Salem**, A. Adinoyi, M. Rahman, H. Yanikomeroglu, and D. Falconer, Radio Resource Management in OFDMA-based Fixed Relay Networks: Literature Review. Deliverable Y1-D1, submitted to Samsung Electronics, Korea, (49 pages), 15 December 2007.
- [13] S. Kuan, "Overview of PCS unlicensed wireless standards in the US," *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, pp. 958-962, October 1996.
- [14] T. Rouse, I. Band, and S. McLaughlin, "Capacity and power investigation of opportunity driven multiple access (ODMA) networks in TDD-CDMA based systems," *IEEE International Conference on Communications (ICC)*, pp. 3202-3206, May 2002.
- [15] H. Zhu and G. Cao, "On improving the performance of IEEE 802.11 with relay-enabled PCF," *Springer Journal on Mobile Networks and Applications*, pp. 423-434, November 2004.
- [16] H. Zhu and G. Cao, "rDCF: A relay-enabled medium access control protocol for wireless ad hoc networks," *IEEE Transactions on Mobile Computing*, pp. 1201-1214, September 2006.
- [17] G. R. Hiertz, S. Max, R. Zhao, D. Denteneer, and L. Berlemann, "Principles of IEEE 802.11s," *International Conference on Computer Communications and Networks*, pp. 1002-1007, August 2007.
- [18] IEEE P802.16j/D1, "Draft IEEE Standard for Local and metropolitan area networks Part 16: Air interface for fixed and mobile broadband wireless access systems: multihop relay specification," pp. 1002-1007, August 2007.
- [19] M. Sternad, T. Svensson, T. Ottosson, A. Ahlen, A. Svensson, and A. Brunstrom, "Towards systems beyond 3G based on adaptive OFDMA transmission," *Proceedings of the IEEE*, 95(12), pp. 2432-2455, December 2007.
- [20] M. Andrews, "Instability of the proportional fair scheduling algorithm for HDR," *IEEE Transactions on Wireless Communications*, pp. 1422-1426, September 2004.
- [21] Y. Ma, "Proportional fair scheduling for downlink OFDMA," *IEEE International Conference on Communications (ICC)*, pp. 4843-4848, June 2007.
- [22] Z. Shen, J. Andrews, and B. Evans, "Adaptive resource allocation in multiuser OFDM systems with proportional rate constraints," *IEEE Transactions on Wireless Communications*, pp. 2726-2737, November 2005.
- [23] C. Y. Huang, H.-H. Juan, M.-S. Lin, and C.-J. Chang, "Radio resource management of

- heterogeneous services in mobile WIMAX systems,” *IEEE Wireless Communications*, pp. 20-26, February 2007.
- [24] S. Chiochan and E. Hossain, “Adaptive radio resource allocation in OFDMA systems: A survey of the state-of-the-art approaches,” *Wireless Communications and Mobile Computing Journal (Wiley)*, 9(4), pp. 513-527, April 2009.
- [25] B. Can, H. Yanikomeroglu, F. Atay Onat, E. Carvalho, and H. Yomo, “Efficient cooperative diversity schemes and radio resource allocation for IEEE 802.16j,” *IEEE Wireless Communications and Networking Conference*, pp. 36-41, April 2008.
- [26] V. Sreng, H. Yanikomeroglu, and D. Falconer, “Capacity enhancement through two-hop relaying in cellular radio systems,” *IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 881-885, March 2002.
- [27] Y. Park and E.-S. Jung, “Resource-aware routing algorithms for multi-hop cellular networks,” *International Conference on Multimedia and Ubiquitous Engineering*, pp. 1164-1167, April 2007.
- [28] X. Qiu and K. Chawla, “On the performance of adaptive modulation in cellular systems,” *IEEE Transactions on Communications*, 47(6), pp. 884-895, June 1999.
- [29] IEEE 802.16, “Mobile multihop relay study group recommendations for the scope and purpose of the mobile multihop relay task group,” 11 November 2005, http://www.ieee802.org/16/sg/mmr/contrib/C80216mmr-05_032.pdf.
- [30] V. Chandrasekhar, J. Andrews, and A. Gatherer, “Femtocell networks: a survey,” *IEEE Communications Magazine*, 46(9), pp. 59-67, September 2008.
- [31] Z. Han and K. Liu, *Resource Allocation for Wireless Networks: Basics, Techniques, and Applications*, Cambridge, 2008.
- [32] S. Kim, X. Wang, and M. Madhian, “Resource allocation in multi-hop OFDMA wireless backhaul networks with cooperative relaying,” *IEEE Global Telecommunications Conference (ICC)*, pp. 5205-5209, November 2007.
- [33] T. Ng and W. Yu, “Joint optimization of relay strategies and resource allocations in cooperative cellular networks,” *IEEE Journal on Selected Areas in Communication*, vol. 25, pp. 328-339, February 2007.
- [34] M. Morelli, C.-C. Kuo, and M.-O. Pun, “Synchronization techniques for orthogonal frequency division multiple access (OFDMA): A tutorial review,” *Proceedings of the IEEE*, 95(7), pp. 1394-1427, July 2007.
- [35] IEEE C802.16m-07/306, “Protocol structure to support cooperative transmission,” *IEEE 802.16 Broadband Wireless Access Working Group*, http://www.ieee802.org/16/tgm/contrib/C80216m-07_306.pdf.
- [36] H. Hu, H. Yanikomeroglu, D. Falconer, and S. Periyalwar, “Range extension without capacity penalty in cellular networks with digital fixed relays,” *IEEE Global Telecommunications Conference (Globecom)*, pp. 3053-3057, November 2004.

- [37] V. Sreng, H. Yanikomeroglu, and D. Falconer, "Relayer selection strategies in cellular networks with peer-to-peer Relaying," *IEEE Vehicular Technology Conference (VTC-Fall)*, pp. 1949-1953, October 2003.
- [38] H. Viswanathan and S. Mukherjee, "Performance of cellular networks with relays and centralized scheduling," *IEEE Transactions on Wireless Communicatios*, 4(5), pp. 2318-2328, September 2005.
- [39] M. Rahman and H. Yanikomeroglu, "QoS provisioning in the absence of ARQ in cellular fixed relay networks through inter-cell coordination," *IEEE Global Telecommunications Conference (Globecom)*, November 2006.
- [40] P. Li, M. Rong, T. Liu, and D. Yu, "Interference modeling and analysis in two-hop cellular network with fixed relays in FDD mode," *International Conference on Wireless Communications, Networking and Mobile Computing (WCNC)*, pp. 427-451, September 2005.
- [41] M. Ahmed, I. Syed, and H. Yanikomeroglu, "On the performance of TDMA-based multi-hop fixed cellular networks with respect to available frequency channels," *IET Communications*, 2(9), pp. 1196-1204, October 2008.
- [42] Y. Liu, R. Hoshyar, X. Yang, and R. Tafazolli, "On the radio resource allocation in enhanced uplink UTRA-FDD with fixed relay stations," *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, pp. 1611-1615, September 2005.
- [43] M. Kaneko and P. Popovski, "Radio resource allocation algorithm for relay-aided cellular OFDMA system," *IEEE International Conference on Communications (ICC)*, pp. 4831-4836, June 2007.
- [44] C. Bae and D.-H. Cho, "Fairness-aware adaptive resource allocation scheme in multi-hop OFDMA Systems," *IEEE Communications Letters*, 11(2), pp. 134-136, February 2007.
- [45] C. Bae and D.-H. Cho, "Adaptive resource allocation based on channel information in multi-hop OFDM systems," *IEEE Vehicular Technology Conference (VTC-Fall)*, September 2006.
- [46] R. Kwak and J. Cioffi, "Resource-Allocation for OFDMA multi-hop relaying downlink systems," *IEEE Global Telecommunications Conference (Globecom)*, pp. 3225-3229, November 2007.
- [47] R. Jain, *The Art of Computer Systems Performance Analysis: Techniques for Experimental Design, Measurement, Simulation and Modeling*, New York: Wiley, 1991.
- [48] J. Lee, S. Park, H. Wang, and D. Hong, "QoS-guarantee transmission scheme selection for OFDMA multi-hop cellular networks," *IEEE International Conference on Communications (ICC)*, pp. 4587-4591, June 2007.
- [49] Ö. Oyman, "Opportunistic scheduling and spectrum reuse in relay-based cellular

- OFDMA networks,” *IEEE Global Telecommunications Conference (Globecom)*, pp. 3699-3703, November 2007.
- [50] M. Kim and H. Lee, “Radio resource management for a two-hop OFDMA relay system in downlink,” *IEEE Symposium on Computers and Communications*, July 2007.
- [51] D. Gesbert S. Kiani, A. Gendhemsjo, and G. Oien, “Adaptation, coordination, and distributed resource allocation in interference-limited wireless networks,” *Proceedings of the IEEE*, 95(12), pp. 2393-2409, December 2007.
- [52] J. Lee, H. Wang, S. Lim, and D. Hong, “A multi-hop user admission algorithm for fixed relay stations with limited capabilities in OFDMA Cellular networks,” *IEEE 18th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, September 2007.
- [53] T. Kim, T.-Y. Min, and C. Kang, “Opportunistic packet scheduling algorithm for load balancing in a multi-hop relay-enhanced cellular OFDMA-TDD System,” *IEEE 14th Asia-Pacific Conference on Communications*, October 2008.
- [54] D. Soldani and S. Dixit, “Wireless Relays for Broadband Access,” *IEEE communications Magazine*, 46(3), pp. 58-66, March 2008.
- [55] M. Neely, E. Modiano, C. Rohrs, “Dynamic power allocation and routing for time-varying wireless networks,” *IEEE Journal on Selected Areas in Communications*, 23(1), pp. 89-103, January 2005.
- [56] L. Tassiulas and A. Ephremides, “Stability properties of constrained queueing systems and scheduling policies for maximum throughput in multihop radio networks,” *IEEE Transactions on Automatic Control*, pp. 1936-1948, December 1992.
- [57] **M. Salem**, A. Adinoyi, M. Rahman, H. Yanikomeroglu, D. Falconer, Y.-D. Kim, W. Shin, and E. Kim, “Fairness-aware radio resource management in downlink OFDMA cellular relay networks,” *IEEE Transactions on Wireless Communications*, 9(5), pp. 1628-1639, May 2010.
- [58] L. Le and E. Hossain, “Multihop cellular networks: Potential gains, research challenges, and a resource allocation framework,” *IEEE Communications Magazine*, 45(9), pp. 66-73, September 2007.
- [59] J. Shi, Z. Zhang, P. Qiu, G. Yu, “Subcarrier and power allocation for OFDMA-based regenerative multi-hop links,” *IEEE Wireless Communications, Networking and Mobile Computing (WCNC)*, pp. 207-210, September 2005.
- [60] D. Niyato and E. Hossain, “A queueing theoretic and optimization-based model for radio resource management in IEEE 802.16 broadband wireless networks,” *IEEE Transactions on Computers*, 55(11), pp. 1473-1488, November 2006.
- [61] H. Wu, C. Qiao, S. De, and O. Tonguz, “Integrated cellular and ad-hoc relay systems: iCAR,” *IEEE Journal on Selected Areas Communications*, vol. 19, pp. 2105-2115, October 2001.

- [62] E. Yanmaz and O. Tonguz, “Dynamic load balancing and sharing performance of integrated wireless networks,” *IEEE Journal on Selected Areas in Communications*, 22(5), pp. 862-872, June 2004.
- [63] E. Yanmaz and O. Tonguz, “The mathematical theory of dynamic load balancing in cellular networks,” *IEEE Transactions on Mobile Computing*, 7(12), pp. 1504-1518, December 2008.
- [64] D. Niyato and E. Hossain, “Integration of IEEE 802.11 WLANs with IEEE 802.16-based multihop infrastructure mesh/relay networks: A game-theoretic approach to radio resource management,” *IEEE Network*, 21(3), pp. 6-14, May-June 2007.
- [65] **M. Salem**, A. Adinoyi, H. Yanikomeroglu, “Integrating self-organizing nomadic relays into OFDMA fixed-relay cellular networks”, submitted to *IEEE Transactions on Mobile Computing*, 19 October 2010.
- [66] Y. Zhao, X. Fang, X. Hu, Z. Zhao, and Y. Long, “Fractional frequency reuse schemes and performance evaluation for OFDMA multi-hop cellular networks,” *International Conference on Testbeds and Research Infrastructures for the Development of Networks & Communities and Workshops*, April 2009.
- [67] W. Lee, M.-V. Nguyen, J. Jeong, B. Keum, and H. Lee, “An Orthogonal Resource Allocation Algorithm to Improve the Performance of OFDMA-Based Cellular Wireless Systems Using Relays,” *IEEE Consumer Communications and Networking Conference*, pp. 917-921, January 2008.
- [68] Y. Song, H. Yang, J. Liu, L. Cai, D. Li, X. Zhu, K. Wu, and H. Liu, “Relay station shared by multiple base stations for inter-cell interference mitigation,” *IEEE 802.16 Broadband Wireless Access Working Group*, IEEE C80216m-08/1436, October 2008.
- [69] M. Rahman and H. Yanikomeroglu, “Interference avoidance through dynamic downlink OFDMA subchannel allocation using intercell coordination,” *IEEE Vehicular Technology Conference (VTC-Spring)*, pp. 1630-1635, May 2008.
- [70] M. Rahman, H. Yanikomeroglu, and W. Wong, “Interference avoidance with dynamic inter-cell coordination for downlink LTE systems,” *IEEE Wireless Communications & Networking Conference (WCNC)*, April 2009.
- [71] A. Goldsmith and S. Chua, “Adaptive coded modulation for fading channels,” *IEEE Transactions on Communications*, vol. 46, pp. 595-602, May 1998.
- [72] H. Zhang, S. Wei, G. Ananthaswamy, D. L. Goeckel, “Adaptive signaling based on statistical characterizations of outdated feedback in wireless communications,” *Proceedings of the IEEE*, 95(12), pp. 2337-2353, December 2007.
- [73] A. Adinoyi and H. Yanikomeroglu, “Cooperative relaying in multi-antenna fixed relay networks,” *IEEE Transactions Wireless Communications*, 6(2), pp. 533-544, February 2007.
- [74] L. Huang, M. Rong, L. Wang, Y. Xue, and E. Schulz, “Resource scheduling for

- OFDMA/TDD based relay enhanced cellular networks,” *IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 1544-1548, March 2007.
- [75] Z. Tang and G. Wei, “Resource allocation with fairness consideration in OFDMA-Based relay networks,” *IEEE Wireless Communications and Networking Conference*, April 2009.
- [76] F. Kelly, A. Maulloo, and D. Tan, “Rate control for communication networks: shadow prices, proportional fairness and stability,” *The Journal of the Operational Research Society*, 49(3), pp. 237-252, March 1998.
- [77] P. Viswanath, D. Tse, and R. Laroia, “Opportunistic beamforming using dumb antennas,” *IEEE Transactions on Information Theory*, 48(6), pp.1277-1294, June 2002.
- [78] W. Anchun, X. Liang, Z. Shidong, X. Xibin, and Y. Yan, “Dynamic resource management in the fourth generation wireless systems,” *International Conference on Communication Technology*, pp. 1095-1098, April 2003.
- [79] L. Xiao and L. Cuthbert, “Improving fairness in relay-based access networks,” *Proceedings of the 11th international symposium on Modeling, analysis and simulation of wireless and mobile systems*, October 2008.
- [80] W.-G. Ahn and H.-M. Kim, “Proportional fair scheduling in relay enhanced cellular OFDMA systems,” *IEEE 19th International Symposium on Personal, Indoor and Mobile Radio Communications*, pp. 1-4, September 2008.
- [81] R. Schoenen, A. Otyakmaz, and B. Walke, “Concurrent operation of half- and full-duplex terminals in future multi-hop FDD based cellular networks,” *IEEE International Conference on Wireless Communications, Networking and Mobile Computing*, October 2008.
- [82] Y. Zhou and H. Sethu, “On the relationship between absolute and relative fairness bounds,” *IEEE Communications Letters*, 6(1), pp. 37-39, January 2002.
- [83] R. Elliott, “A measure of fairness of service for scheduling algorithms in multiuser systems,” *IEEE Canadian Conference on Electrical and Computer Engineering*, vol. 3, pp. 1583-1588, May 2002.
- [84] A.V. Babu, L. Jacob, “Fairness analysis of IEEE 802.11 multirate wireless LANs,” *IEEE Transactions on Vehicular Technology*, 56(5), part 2, pp. 3073-3088, September 2007.
- [85] F. Bokhari, H. Yanikomeroglu, W. K. Wong, and M. Rahman, “Fairness assessment of the adaptive token bank fair queuing scheduling algorithm,” *IEEE Vehicular Technology Conference (VTCFall)*, September 2008.
- [86] J. Vicario, A. Bel, A. Morell, and G. Seco-Granados, “Outage probability versus fairness trade-off in opportunistic relay selection with outdated CSI,” *EURASIP Journal on Wireless Communications and Networking*, article ID 412837, January 2009.
- [87] IEEE 802.16m-08/004r5, “Evaluation Methodology Document

- (EMD),” *IEEE 802.16 Broadband Wireless Access Working Group*, http://www.ieee802.org/16/tgm/core.html#08_004. January 2009.
- [88] M. Kobayashi and G. Caire, “Joint beamforming and scheduling for a multi-antenna downlink with imperfect transmitter channel knowledge,” *IEEE Journal on Selected Areas in Communications*, 25(7), pp. 1468-1477, September 2007.
- [89] P. Parag, S. Bhashyam, and R. Aravind, “A subcarrier allocation algorithm for OFDMA using buffer and channel state information,” *Vehicular Technology Conference (VTC-Fall)*, pp. 622-625, September 2005.
- [90] **M. Salem**, A. Adinoyi, H. Yanikomeroglu, and Y.-D. Kim, “Fair resource allocation towards ubiquitous coverage in OFDMA-based cellular relay networks with asymmetric traffic,” submitted to *IEEE Transactions on Vehicular Technology*, July 2010, revised October 2010.
- [91] T. S. Rappaport, *Wireless Communications: Principle and Practice*, 2nd edition, Prentice Hall, 2001.
- [92] WINNER, “WINNER II Channel Models,” available online: <https://www.ist-winner.org/WINNER2-Deliverables/D1.1.2v1.1.pdf>, March 2008.
- [93] H. Kim and Y. Han, “A proportional fair scheduling for multicarrier transmission systems,” *IEEE Communications Letters*, 9(3), pp. 210-212, March 2005.
- [94] A. Eryilmaz and R. Srikant, “Fair resource allocation in wireless networks using queue-length-based scheduling and congestion control,” *IEEE/ACM Transactions on Networking*, 15(6), pp. 1333-1344, December 2007.
- [95] M. Neely, E. Modiano, and C. Rohrs, “Power and server allocation in a multi-beam satellite with time varying channels,” *IEEE INFOCOM*, New York, pp. 1451-1460, June 2002.
- [96] G. Kramer, M. Gastpar, and P. Gupta, “Cooperative strategies and capacity theorems for relay networks,” *IEEE Transactions on Information Theory*, 51(9), pp. 3037-3063, September 2005.
- [97] A. Høst-Madsen and J. Zhang, “Capacity bounds and power allocation for wireless relay channels,” *IEEE Transactions on Information Theory*, 51(6), pp. 2020-2040, June 2005.
- [98] **M. Salem**, A. Adinoyi, M. Rahman, H. Yanikomeroglu, D. Falconer, and Y.-D. Kim, Apparatus and Method for Allocating Subchannels and Controlling Interference in OFDMA Systems. Filed by Samsung, Korea patent application no: P2008-0054726 (filing date: 11 June 2008), US patent application no: 12/341,933 (22 December 2008), international patent application no: PCT/KR2009/002119 (23 April 2009).
- [99] **M. Salem**, A. Adinoyi, M. Rahman, H. Yanikomeroglu, D. Falconer, Y.-D. Kim, W. Shin, and E. Kim, “Fairness-aware joint routing and scheduling in OFDMA-based cellular fixed relay networks,” *IEEE International Conference on Communications*,

June 2009.

- [100] **M. Salem**, A. Adinoyi, M. Rahman, H. Yanikomeroglu, and D. Falconer, Radio Resource Management through Joint Routing and Fair Scheduling and Multi-cell Coordination in OFDMA Fixed Relay Networks. Deliverable Y1-D2, submitted to Samsung Electronics, Korea, (50 pages), 15 March 2008.
- [101] H. W. Kuhn, "The Hungarian method for the assignment problem," *Naval Research Logistic Quarterly*, 2(1), pp. 83-97, 1955.
- [102] H. Yin and H. Liu, "An efficient multiuser loading algorithm for OFDM-based broadband wireless systems," *IEEE Global Telecommunications Conference*, pp. 103-107, December 2000.
- [103] E. Dahlman, H. Ekstrom, A. Furuskar, Y. Jading, J. Karlsson, M. Lundevall, and S. Parkvall, "The 3G long-term evolution-Radio interface concepts and performance evaluation," *IEEE Vehicular Technology Conference (VTC-Spring)*, pp. 137-141, May 2006.
- [104] M. Moretti and A. Todini, "A resource allocator for the uplink of multi-cell OFDMA systems," *IEEE Transaction on Wireless Communications*, 6(8), pp. 2807-2812, August 2007.
- [105] K. Ramadas and R. Jain, "Mobile WiMAX Part I: A technical overview and performance evaluation," *WiMAX Forum*, September 2007. wimaxforum.org/technology/downloads/Mobile_WiMAX_Part1_Overview_and_Performance.pdf.
- [106] Ö. Oyman, "Opportunistic scheduling and spectrum reuse in relay-based cellular networks," *IEEE Transactions on Wireless Communications*, 9(3), pp. 1074-1085, March 2010.
- [107] **M. Salem**, A. Adinoyi, H. Yanikomeroglu, D. Falconer, and Y.-D. Kim, Method for Performing Fair Resource Allocation in OFDMA-based Relay-Networks. Filed by Samsung, Korea, patent application no: P2009-0022132 (filing date: 16 March 2009), USA patent application no: 12/567,776 (September 2009), and international application is underway.
- [108] **M. Salem**, A. Adinoyi, H. Yanikomeroglu, D. Falconer, and Y.-D. Kim, "A fair radio resource allocation scheme for ubiquitous high-data-rate coverage in OFDMA-based cellular relay networks," *IEEE Global Communications Conference*, December 2009.
- [109] IEEE 802.16m-09/0034r3, "System Description Document (SDD)," *IEEE 802.16 Broadband Wireless Access Working Group*, http://www.ieee802.org/16/tgm/core.html#08_004. July 2010.
- [110] D. Niyato and E. Hossain, "Service differentiation in broadband wireless access networks with scheduling and connection admission control: A unified analysis," *IEEE Transactions on Wireless Communications*, 5(12), pp. 293-301, December 2006.

- [111] M. Mehrjoo, M. Awad, M. Dianati, and X. S. Shen, "Design of fair weights for heterogeneous traffic scheduling in multichannel wireless networks," *IEEE Transactions on Communications*, 58(10), pp. 2892-2902, October 2010
- [112] 3GPP Technical Report TR 36.942 "Evolved universal terrestrial radio access (E-UTRA); Radio frequency (RF) system scenarios," available online: <http://www.3gpp.org/ftp/specs/html-info/36942.htm>, April 2010.
- [113] C. Prehofer, C. Bettstetter, "Self-organization in communication networks: principles and design paradigms," *IEEE Communications Magazine*, 43(7), pp. 78-85, July 2005.
- [114] SOCRATES, "Self-optimisation and self-configuration in wireless networks," *European Research Project*, January 2008 - December 2010. <http://www.fp7-socrates.eu.org>.
- [115] INFSO-ICT-216284 SOCRATES, "Use cases for self-organizing networks," Deliverable D2.1, Version 1.0, March 2008. Available at <http://www.fp7-socrates.org/files/Publications/>
- [116] 3GPP, "Self-configuring and self-optimizing network use cases and solutions," Technical Report TR 36.902. Available at <http://www.3gpp.org>.
- [117] INFSO-ICT-216284 SOCRATES, "Framework for the development of self-organisation methods," Deliverable D2.4, Version 1.0.3, July 2008. Available at <http://www.fp7-socrates.org/files/Publications/>
- [118] E. Yanmaz, O. Tonguz and S. Dixit, "Self-organization in cellular wireless networks via fixed relay nodes," *IEEE Global Telecommunications Conference (GLOBECOM)*, pp. 1-5, November 2006.
- [119] C. Koh, K. Yoon, K. Sohn, S. Moon and Y. Kim, "Self-organized spatial reuse scheduling in multihop cellular systems," *IEEE 19th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, September 2008.
- [120] P. Jiang, J. Bigham and J. Wu, "Self-organizing relay stations in relay based cellular networks," *Computer Communications*, 31(13), pp. 2937-2945, August 2008.
- [121] L. Xu, Y. Chen and Y. Gao, "Self-organizing load balancing for relay based cellular networks," *IEEE International Conference on Computer and Information Technology (CIT)*, pp. 791-796, July 2010.
- [122] M. Salem, A. Adinoyi, H. Yanikomeroglu, and Y.-D. Kim, "Radio resource management in OFDMA-based cellular networks enhanced with fixed and nomadic relays", *IEEE Wireless Communications and Networking Conference*, April 2010.
- [123] M. Salem, A. Adinoyi, H. Yanikomeroglu, and Y.-D. Kim, "Nomadic relay-directed joint power and subchannel allocation in OFDMA-based cellular fixed relay networks", *IEEE Vehicular Technology Conference (VTC-Spring)*, May 2010.
- [124] **M. Salem**, A. Adinoyi, H. Yanikomeroglu, and Y.-D. Kim, Method for Relaying Data in Wireless Network and Personal Relay of Enabling the Method, and Mobile

- Device for Communicating with the Personal Relay. Filed by Samsung, Korea; Korea patent application no: P2009-0084026 (application date: 07 September 2009).
- [125] **M. Salem**, A. Adinoyi, H. Yanikomeroglu, and D. Falconer, Distributed Radio Resource Management in OFDMA-based Multi-cellular Networks Enhanced with Fixed and Nomadic Relays. Deliverable Y1-D3, submitted to Samsung Electronics, Korea, (32 pages), 15 September 2008.
- [126] **M. Salem**, A. Adinoyi, H. Yanikomeroglu, and D. Falconer, Distributed radio resource management in OFDMA-based multi-cellular networks enhanced with fixed and nomadic relays. Deliverable Y2-D1, submitted to Samsung Electronics, Korea, (59 pages), 15 January 2009.
- [127] 3GPP Release 10, "A description document - Overview of 3GPP Release 10; summary of all Release 10 features," June 2010.
Available online: <http://www.3gpp.org/Release-10>.
- [128] S. Szyszkowicz, H. Yanikomeroglu and J. Thompson, "On the feasibility of wireless shadowing correlation models," *IEEE Transactions on Vehicular Technology*, 59(9), pp. 4222-4236, November 2010.
- [129] D. Niyato, E. Hossain, D. Kim, and Z. Han, "Relay-centric radio resource management and network planning in IEEE 802.16j mobile multihop relay networks," *IEEE Transactions on Wireless Communications*, 8(12), pp. 6115-6125, December 2009.
- [130] B. Babadi and V. Tarokh, "A distributed asynchronous algorithm for spectrum sharing in wireless ad hoc networks," *Annual Conference on Information Sciences and Systems (CISS)*, pp. 831-835, March 2008.
- [131] J. Munkers, "Algorithms for the assignment and transportation problems," *Journal of the Society for Industrial and Applied Mathematics*, pp. 3238, 1957.
- [132] J. Zander, "Distributed cochannel interference control in cellular radio systems," *IEEE Transactions on Vehicular Technology*, 41(3), pp. 305-311, August 1992.
- [133] S. Ulukus and R. Yates, "Stochastic power control for cellular radio systems," *IEEE Transactions on Communications*, 46(6), pp. 784-798, June 1998.
- [134] M. Pischella and J.-C. Belfiore, "Power control in distributed cooperative OFDMA cellular networks," *IEEE Transactions on Wireless Communications*, 7(5), pp. 1900-1906, May 2008.
- [135] GreenNet 2010, Green Wireless Communications and Networks Workshop, co-located with the *IEEE Vehicular Technology Conference (VTC-Fall)*, September 2010. www.ieeevtc.org/vtc2010fall/greenet.pdf
- [136] Online article, "BARCELONA MWC 2009: The restructuring of a sector," *Mobile World Congress*, Barcelona, March 2009. Available at:
<http://socioadellainformacion.telefonica.es/jsp/articulos/>
- [137] C. Wong, R. Cheng, K. Letaief, and R. Murch, "Multiuser OFDM with adaptive

- subcarrier, bit, and power allocation,” *IEEE Journal on Selected Areas in Communications*, 17(10), pp. 1747-1758, October 1999.
- [138] F. Chau, “Mobile goes green: reducing power consumption is just the first step,” online article, *Telecom Asia*, January 2008. Available at:
http://findarticles.com/p/articles/mi_m0FGI/is_1_19/ai_n24381287/
- [139] **M. Salem**, A. Adinoyi, H. Yanikomeroglu, and D. Falconer, Distributed radio resource management with Power Control in OFDMA-based multi-cellular networks enhanced with fixed and nomadic relays. Deliverable Y2-D2, submitted to Samsung Electronics, Korea, (34 pages), 15 April 2009.
- [140] S. Al-Ahmadi and H. Yanikomeroglu, “On the approximation of the generalized-K distribution by a Gamma distribution for modeling composite fading channels”, *IEEE Transaction on Wireless Communications*, 9(2), pp. 706-713, February 2010.
- [141] Z. Han, T. Himsoon, W. Siriwongpairat, and K. Liu, “Resource allocation for multiuser cooperative OFDM networks: Who helps whom and how to cooperate,” *IEEE Transactions on Vehicular Technology*, 58(5), pp. 2378-2391, June 2009.