

© 2010

Chandrasekharan Raman

ALL RIGHTS RESERVED

**RELAYING AND SCHEDULING IN INTERFERENCE  
LIMITED WIRELESS NETWORKS**

by

**CHANDRASEKHARAN RAMAN**

A Dissertation submitted to the  
Graduate School—New Brunswick  
Rutgers, The State University of New Jersey  
in partial fulfillment of the requirements

for the degree of

**Doctor of Philosophy**

**Graduate Program in Electrical and Computer Engineering**

written under the direction of

**Prof. Roy D. Yates,**

**Prof. Narayan B. Mandayam and Dr. Gerard J. Foschini**

and approved by

---

---

---

---

---

**New Brunswick, New Jersey**

**May, 2010**

## ABSTRACT OF THE DISSERTATION

# Relaying and Scheduling in Interference Limited Wireless Networks

By Chandrasekharan Raman

Dissertation Directors: Prof. Roy D. Yates,

Prof. Narayan B. Mandayam and Dr. Gerard J. Foschini

In this dissertation, we address two issues related to communication in interference-limited wireless networks. In the first part of the thesis, we study benefits of deploying inexpensive half-duplex relays in interference-limited cellular system. We study two relaying schemes in a downlink system, users sharing the same frequency band. In the first collaborative relaying scheme, relays help users after they decode the message intended for the user by a collaborative power addition scheme. We evaluate power savings and rate improvement for delivering a common rate for 90% of users in the system. We then consider power control at the base stations and relays in order to manage the interference. In both cases, the ability of relays to reduce the peak base station transmit power while delivering the baseline rate or alternately to increase the user rate is computed. When power control is employed, the peak power saving is 2.6 dB and the average total power in the system can be reduced by 3 dB. We also observe a 34% improvement over the baseline in the users common rate. In the second orthogonal relaying scheme, we study a simple scheme where the base station and relays transmit in orthogonal time slots. We find that the performance of the simple orthogonal relaying scheme comes very close to that of the collaborative power addition scheme.

In the second part of the thesis, we consider a centralized spectrum server that coordinates the transmissions of a group of links sharing a common spectrum. Links employ on-off modulation with fixed transmit power when active. The rate obtained by an active link depends on the activity of all other links. With knowledge of the link gains in the network, the spectrum server schedules the on/off periods of the links so as to satisfy constraints on link fairness and efficiency. We then extend the centralized scheduling framework to multi-hop wireless networks with interfering links. A framework for cross-layer scheduling of end-to-end flows in a multi-hop wireless network with links sharing a common spectrum is presented. Given a set of known routes, the optimization framework can be used to find the flow on each route in order to maximize an objective function. The framework encompasses a variety of physical layer transmission schemes. The rates in the individual links are determined based on the transmission strategy employed. With the knowledge of the link rates in the network, the spectrum server schedules the rates on the links and flows on the sessions to maximize a utility function of the source rates. The schedules are a collection of time shared transmission modes (sets of active links).

## Acknowledgements

*na choraharyam na raajaharyam na bhraatrubhaajyam na cha bhaarakari  
vyaye krute vardhata eva nityam vidyaadhanam sarvadhana pradhaanam*

– an Ancient Indian adage.

*Not stolen by thieves, not seized by kings, not divided amongst brothers, not heavy to carry. The more you spend, the more it flourishes always — the wealth of knowledge is the most important among all kinds of wealth.*

Looking back into the past five years, I am overwhelmed by the great wealth of knowledge I have gained. I am extremely thankful to my advisors, Professors Roy Yates, Narayan Mandayam and Gerard Foschini for guiding me in this pursuit. They will always be great inspiration for me to set high standards in my work.

It has been an enjoyable experience to work with Roy. He never missed an opportunity to give good advice. Roy taught the approach of solving any research problem by addressing the simplest non-trivial problem. His child-like enthusiasm to any research idea is infectious. Many a times, I wonder how he asks the right questions in research seminars. Narayan was a great mentor and taught by living as an example. I thank him for bringing me to WINLAB and for his belief in my abilities, especially during the times when I doubted them myself. I also benefited a lot by taking his courses. Working with Jerry Foschini at Bell Labs has been an invaluable experience. His meticulous approach improved my writing and presentation skills a lot. He instilled in me the habit of solving a problem from scratch.

During the course of my PhD, I was fortunate to interact with some great people, who are giants in the field of wireless communications. At WINLAB, Larry Greenstein, Dick Frankiel, Rich Howard, Giovanni Vannucci gave me valuable advice. Thanks to Prof. Chris Rose, who was always open for discussion and to have served on my

thesis committee. Reinaldo Valenzuela was kind enough to host me at Bell Labs. I learned a lot during my visit to Crawford Hill labs through my interaction with many researchers at Bell Labs. Thanks to Harish Viswanathan and Sivarama Venkatesan for their insights into the relay problem. My internship at Qualcomm, San Diego was a memorable experience. I thank Ashwin Sampath to have introduced the relay problem that fructified as a chapter in my thesis. It was a pleasure interacting with Jack Holtzman and other colleagues at Qualcomm.

WINLAB is a great place for intellectual growth. The diverse areas of research, easily accessible facilities, educative seminar series, helpful administrative staff make it one of the best places to learn wireless communications and systems. I thank Professors Dipankar Raychaudhuri, Predrag Spasojevic and Ivan Seskar for giving me the opportunity to work on projects other than my thesis topics. Students are the greatest asset of WINLAB. I benefited a lot, academically and otherwise, by personal interaction with WINLAB students. I thank my colleagues Jasvinder, Hithesh, Joydeep, Lalitha, Rahul, Sanjit, Suhas, Özge, Kemal, Ivana, Michelle, Silvija, Omer, Liang, Zang, Dan, Song and many others for the discussion on interesting research problems and their help in my research. I also thank Kishore, Sachin, Haris, Gayathri, Gautam and Mehsut for the good academic and non-academic discussions. My special thanks go to all those who commuted to WINLAB with me; they kept me motivated to go to work without a break!

I am grateful to the wonderful set of friends outside school, to have alleviated the stress from graduate life. I will always cherish the long phone calls with SK, Girish and others. Easwar was always there when I needed someone to talk about my problems. My association with Asha for Education helped me understand all the privileges we take for granted and also won me many good friends.

I cannot thank my sister and brother enough for their support. They took care of most of the issues back home, most of which I was not even aware of. Finally, I owe my debt of gratitude to my parents for their unconditional love and faith through this journey. Even though they lived miles away, they had more concern than me for the successful completion of my PhD. To them, I dedicate this work.

# Dedication

*To my Parents*

and

*Teachers*

for giving me the greatest gift  
of Education

# Table of Contents

<b>Abstract</b> . . . . .	ii
<b>Acknowledgements</b> . . . . .	iv
<b>Dedication</b> . . . . .	vi
<b>List of Tables</b> . . . . .	xi
<b>List of Figures</b> . . . . .	xii
<b>1. Introduction</b> . . . . .	1
<b>2. Relaying in Downlink Cellular Systems</b> . . . . .	4
2.1. Motivation and background . . . . .	4
Throughput gains due to relay deployments . . . . .	4
Relay placement . . . . .	5
Lack of good models for relaying in cellular systems . . . . .	6
Fairness . . . . .	7
2.2. Related work and overview of our contribution . . . . .	8
2.3. The set-up . . . . .	10
2.3.1. Placement of users and relays in the system . . . . .	13
2.4. Organization . . . . .	14
2.5. Collaborative relaying in cellular networks (CPA) . . . . .	16
2.6. CPA with peak power transmissions (P-CPA) . . . . .	18
2.6.1. Principle of operation . . . . .	18
P-CPA Baseline . . . . .	18
P-CPA system with Relays . . . . .	19
2.6.2. User discarding methodology . . . . .	21

2.6.3.	Network Operation and Simulation Aspects . . . . .	23
2.6.4.	Simulation results . . . . .	24
	Power savings . . . . .	24
	Rate gains . . . . .	24
2.7.	Power control based collaborative relaying (PC-CPA) . . . . .	26
2.7.1.	Principle of operation . . . . .	27
	PC-CPA Baseline . . . . .	27
	PC-CPA System with Relays . . . . .	28
2.7.2.	Optimization framework . . . . .	29
	Minimizing the total instantaneous transmit powers . . . . .	29
	Minimizing the peak transmit power . . . . .	32
	Improving the common rate . . . . .	32
2.7.3.	User discarding methodology . . . . .	33
2.7.4.	Network Operation and Simulation Aspects . . . . .	35
	Baseline operation . . . . .	35
	PC-CPA with relays: Average power savings . . . . .	35
	PC-CPA with relays: Peak power savings . . . . .	36
2.7.5.	Simulation Results . . . . .	36
	Power savings . . . . .	36
	Rate gains . . . . .	36
2.8.	Orthogonal Relaying . . . . .	37
2.8.1.	Network Operation and Simulation Aspects . . . . .	38
	Baseline . . . . .	38
	System with relays . . . . .	38
2.8.2.	User discarding method . . . . .	39
	Baseline . . . . .	39
	System with relays . . . . .	39
2.8.3.	Simulation Results . . . . .	40
2.9.	Conclusions . . . . .	42

<b>3. Fair and Efficient Scheduling of Variable Rate Links via a Spectrum Server</b>	<b>43</b>
3.1. Introduction	43
3.1.1. Related Work	45
3.2. System Model	46
Gaussian Interference Channel	49
Gaussian Multiple Access Channel	50
3.3. Maximum Sum Rate Scheduling	53
3.3.1. No minimum rate constraint	54
3.3.2. Non-zero minimum rate constraint	55
3.3.3. Simulation Results	57
3.3.4. Maximum sum rate schedule with high SNR links	59
3.4. Fair scheduling	62
3.4.1. Equal time scheduling	65
3.5. Energy Efficient scheduling	66
3.5.1. Simulation Results	70
3.6. A unified framework for centralized scheduling	72
3.7. Towards a distributed algorithm - Random scheduling	73
3.8. Conclusion	75
3.9. Appendix	76
3.9.1. Case (i): $\mathcal{T}_1^*$ is non-empty	77
3.9.2. Case (ii): $\mathcal{T}_1^*$ is empty	77
<b>4. Cross Layer Scheduling in Multihop Networks via a Spectrum Server</b>	<b>79</b>
4.1. Cross-layer Resource Allocation of End-to-End Flows	79
4.1.1. Related Work	81
4.2. System Model	82
4.2.1. Physical Layer and Interference Model	82
4.2.2. Traffic flow model	84

4.3. Simulation Results . . . . .	87
4.3.1. Maximizing sum of session rates in the network . . . . .	88
4.3.2. Performance with successive interference cancellation at the receiver	89
4.3.3. Relation to Toumpis's work . . . . .	91
4.4. Conclusion . . . . .	96
<b>5. Conclusion</b> . . . . .	<b>98</b>
Propagation models for cellular relay networks . . . . .	98
Practical coding schemes for implementing collaborative relaying	98
Low complexity schemes to find optimal order of base transmission	99
<b>References</b> . . . . .	<b>100</b>
<b>Curriculum Vitae</b> . . . . .	<b>105</b>

## List of Tables

2.1. Parameters used in simulations . . . . .	13
2.2. PERFORMANCE EVALUATION OF RELAYS IN DOWNLINK CELLULAR SYSTEMS . . . . .	15
2.3. P-CPA relaying. Base station and relays transmit at peak power . . . . .	26
2.4. PC-CPA based relaying. Base station and relays employ power control . . . . .	37
2.5. SUMMARY OF GAINS DUE TO RELAYING: 19 cells, 57 sector network, one user per sector . . . . .	41
4.1. Schedules for the transmission modes for different inter-node distances when single user decoding is used at the receiver . . . . .	90
4.2. Schedules for the transmission modes for different inter-node distances when successive decoding (stronger link decoded first) is employed at the receiver . . . . .	92

## List of Figures

2.1. Wrap-around simulation model. The center ring of 19 cells are used for the simulation. The surrounding cell activity is mirrored in the center ring. The direction of the arrows represent the direction of the main lobe of the sectorized antenna. . . . .	10
2.2. Antenna gain pattern (from [1]) as a function of the horizontal angle in degrees. The mathematical expression for the gain is given in equation (2.1). . . . .	11
2.3. Position of relay location in a cell. The relays (represented by small circles) are placed at half the cell radius in the direction (given by the arrows) of the main lobe of the sector antenna. The base station at the center of the cell is represented by a square. . . . .	12
2.4. Collaborative relaying: before relay decodes the message . . . . .	17
2.5. Collaborative relaying: after relay decodes the message . . . . .	17
2.6. Variation of outage with relay powers and base station powers for the P-CPA scheme. As we increase the base station powers with no relays in the system (ratio = 0), the outage decreases and saturates at around 5%, due to the interference limit. The interference limit sets in very quickly even for smaller values of relay powers. . . . .	25
3.1. Graph of network showing the nodes and directed links . . . . .	47
3.2. Graph of network showing transmission mode corresponding to [1 0 1 0 0]	48
3.3. Multiaccess Channel . . . . .	50
3.4. Rate region of the multiaccess channel . . . . .	51
3.5. A set of source-destination pairs . . . . .	57
3.6. Variation of sum rates and individual rate as a function of $\mathbf{r}_{\min}$ . . . . .	58

3.7. Illustration of the solidarity property. The maximizing the common rate leads to max-min fairness in the Gaussian interference model. Whereas in the multiple access model, maximum common rate and max-min fair rates are unequal. . . . .	63
3.8. Comparison between rates of the links under different settings . . . . .	64
3.9. Set of five links each of length $d = 10$ . . . . .	69
3.10. Variation of sum rate with efficiency $\epsilon$ . . . . .	70
3.11. Variation of rates of individual links with efficiency $\epsilon$ for $\mathbf{r}_{\min} = 0.5$ . . .	71
3.12. Variation of rates of individual links with efficiency $\epsilon$ for $\mathbf{r}_{\min} = 0.5$ . . .	72
3.13. Comparison of the average rates obtained by equal time scheduling and random probabilistic scheduling. . . . .	74
4.1. An example of a simple multihop network. The solid arrows are instances of links in the network and dotted lines show a flow through the network	80
4.2. Graph of a network: (a) Nodes and the directed links (b) Illustration of transmission mode $\mathbf{t}=[1\ 0\ 1\ 0\ 0]$ in which links 1 and 3 are active . . . .	83
4.3. Graph of network showing the nodes and directed links. The session is split into two flows $f_{11}$ and $f_{21}$ . $f_{11}$ flows through links (1, 2) and (2, 3) while $f_{21}$ goes through (1, 3). . . . .	85
4.4. Linear network of 5 nodes, with inter-node distance $d$ . The network has 10 links. A session originating at node 1 and terminating at node 5 is shown. The session can take 8 distinct routes in the network. . . . .	87
4.5. Variation of session rates as the inter-node distance $d$ increases. Single hop routing is optimal for smaller values of $d$ and the four-hop route is asymptotically optimal for large values of $d$ . . . . .	89
4.6. Comparison of throughput gains due to successive interference cancellation at the destination node . . . . .	91
4.7. A four node linear network. We consider two flows sharing the available resources. . . . .	93

# Chapter 1

## Introduction

This thesis deals with communication over interference limited wireless networks. Interference is an inevitable ingredient in multi-user wireless communication systems. Wireless systems have been designed to manage interference in many ways. At one end, orthogonalizing multiple transmissions in time and/or frequency simplifies the problem to one of multiple simultaneous point-to-point communication links. Examples for such transmission schemes are Time Division Multiple Access (TDMA) and Frequency Division Multiple Access (FDMA) [2, Chapter 15]. At the other extreme, spreading the signal and interference over a wide band can make the interference look like additive Gaussian noise and the problem again simplifies to a point-to-point communication link. An example of such a scheme is the Code Division Multiple Access (CDMA) [3]. More advanced techniques such as multiuser detection lead to complexity in the receiver structures in practical systems [4]. In fact, the best (capacity-achieving) transmission scheme for the interference channel is still unknown [5].

A practical example of an interference limited wireless network is the CDMA cellular network. In the early days of cellular system, voice was the predominant type of traffic carried by the system. Power control in CDMA systems proved to be an effective way to handle interference when a certain voice quality (in terms of minimum signal to interference plus noise ratio) was to be guaranteed to the users in the system. Of late, when data traffic has become significant, maximizing end-to-end throughput under strict quality of service constraints has turned out to be a very important criterion for good system design. Recently, there have been many proposals to improve the throughput of cellular systems. One such proposal is to use dedicated relays in cellular networks.

Though relays can improve the coverage and throughput in cellular systems, uncoordinated transmissions from additional relays might degrade the performance of the system by increasing the interference in the system. In the first part of the thesis, we study a deployment of half-duplex relays, in which a relay cannot simultaneously transmit and receive in the same band, in cellular systems. We find that interference limits the throughput performance as well as the power saving benefits that can be accrued in a downlink cellular system. We propose power control in the presence of relays in cellular systems and evaluate the power savings in the system. We evaluate two different relaying methodologies — a collaborative scheme where the relays help the base station in delivering the message to the destination and an orthogonal a simple scheme where the base station and relay transmit in orthogonal time slots. An interesting observation is that the performance of both relaying schemes come very close to each other, indicating that simpler schemes buy us most of the gains due to relaying.

In the second part of the thesis, we consider a scenario in emerging wireless systems where radios conforming to multiple competing technologies coexist in a geographical region interfering with each other. The emerging cognitive radio networks fall into this category. In such a scenario, the problem of interference can be more severe since the various transmitter/receiver link pairs operate independently, each according to its own protocol specifications. Interference aware scheduling can enhance the performance of these systems. We propose a centralized framework to schedule these heterogeneous radios. We assume that efficient open access to spectrum can be resolved by impartial “spectrum servers” that can obtain information about the interference environment through measurements contributed by different terminals, and then offer suggestions for efficient coordination to interested service subscribers. As observed in [6], likely neighborhood information could include various levels of time and frequency utilization, descriptions of nodes in a neighborhood, and potentially, spatial positions as well. The role of such a spectrum server for wireless network coordination is reminiscent of the role of the DHCP (Dynamic Host Configuration Protocol) server in the coordination problem that arises among nodes in the Internet.

Specifically, we develop a centralized link scheduling framework to study the performance bounds on interference limited wireless networks. Links employ ON-OFF modulation with fixed transmit power when active. In the ON state, a link obtains a data rate that is determined by the physical layer strategy employed at the receiving node of the link. With knowledge of the link gains in the network, the spectrum server schedules the ON/OFF periods of the links so as to satisfy constraints on link fairness and efficiency. Using a graph theoretic model for the network and a linear programming formulation, we express fairness constraints as lower bounds on the average minimum rate for each link. Efficiency constraints are expressed as lower bounds on the ratio of the average rate to the average transmit power for each link. Subject to fairness and efficiency constraints, the spectrum server finds a schedule that maximizes the average sum rate. The resulting schedule is a collection of time shared transmission modes (sets of active links). The transmission modes are reminiscent of spatial reuse patterns in cellular systems.

We then extend the above model to multihop wireless networks. We use a multi-commodity flow model for routing various flows and the above interference model for the physical layer. We obtain some intuition to which links should be scheduled to optimize various objective functions. The objective functions captures the throughput, efficiency and fairness requirements. The solution to the optimization problem gives the corresponding operating point in the rate region of the set of links in the system.

## Chapter 2

# Relaying in Downlink Cellular Systems

### 2.1 Motivation and background

The deployment of relays in cellular system has recently been standardized in the WiMAX, IEEE 802.16j standard and is a topic of discussion in the advanced specifications of 3GPP-Long Term Evolution (LTE) [7]. Although commercial relay deployments in cellular systems are not prominent at present, future wireless cellular systems will involve operation with dedicated relays for improving coverage, increasing cell-edge throughput and delivering high data rates. The proposed architecture is such that relays would be placed at certain locations (planned or unplanned) in the cell to help in forwarding the message from the base station to the user in the downlink, and from the user to the base station in the uplink. Relays will be more sophisticated than simple repeaters and could perform some digital base band processing to help the destination terminal get better reception. Practical constraints on the radio demand that the relays are half-duplex and hence cannot simultaneously transmit and receive in the same band. These relays will rely on air interfaces, and hence avoid the considerable backhaul costs involving data aggregation and infrastructure costs associated with backbone connectivity. However, there are a lot of open issues that require research to answer. We present some of these issues in the sequel:

#### 1. Throughput gains due to relay deployments

In cellular networks that are coverage limited, deploying relays can help in multi-hop transmission and provide power gains due to reduction of distance attenuation [8]. These power gains, in turn, translate to throughput improvements for the

power limited edge users. However in interference limited settings, as is common in cellular systems, uncoordinated transmission by relays lead to increase of the overall interference levels in the cell and could be counter-productive by reducing the signal-to-interference plus noise (SINR) levels of users in the system. Coordination of transmissions in the system would require centralized control and incur high costs and overhead, especially in the uplink. Resource sharing across the backhaul and access links becomes important to achieve gains in such scenarios. Thus, there is a need for a thorough evaluation of throughput improvements in a cellular system. In the cellular systems literature, there have been simulation studies to evaluate throughput gains in cellular systems, e.g., [9–11]. Even though the studies were conducted under different sets of (idealized) assumptions, throughput improvements in interference limited cellular systems are shown to be around 30% to 40% for the edge users. In this dissertation, we evaluate the gains due to relay deployment by two different relaying strategies and the results indicate that throughput gains are of the same order. However, there may exist better practical schemes — which remain open — or specific scenarios where relays provide larger throughput improvements. A simple case where relays provide throughput improvements is the downlink scenario where the edge user is in a deep canyon and the relay is placed in the line of sight of both the base station and the shadowed user [10].

## 2. Relay placement

The benefits from relay deployment depend on where the relays are placed in the cell. Throughput improvements depend on the transmit power, relay antenna pattern and location of the relays in the system. Placing relays closer to an edge user may help the edge user. However, when relay transmissions are uncoordinated, the relays may cause near line-of-sight interference to an edge user of the neighboring cell. The optimal relay placement depends on the transmission and scheduling strategies, transmit power of the relays etc. A closely related issue to

the relay placement problem is the choice of height of the deployed relays. In macro-cellular environments, propagation characteristics of the base-relay link and the relay-user link could be completely different, depending on whether the relays are mounted on tall poles or on roof tops. These factors may very well affect the system performance due to relay deployments. There are not many measurement based models to cover all the scenarios of relay placement; some empirical models are described in [1]. These issues apart, service providers often do not have much choice in placing the relays in a given geographical area.

### 3. Lack of good models for relaying in cellular systems

Multihopping in wireless networks has been studied in the context of ad hoc networks and peer-to-peer networks [12]. The main issue addressed in such networks is the routing problem. Interference constraints are abstracted as combinatorial constraints and many insightful results and good algorithms have been proposed to improve the throughput of such networks. Cellular networks, however, are unique in that the traffic is one-to-many in the downlink and many-to-one in the uplink. Direct application of the solutions obtained in the context of ad hoc networks are not optimal for cellular systems. Hence, performance evaluation of relays in cellular system requires fresh thinking into the problem.

On the other hand, the information theoretic relay channel [13] has been an active area of research for three decades now. But for some coding strategies proposed by Cover and El Gamal in [14] for special cases of the single relay channel, the capacity of the general relay channel is still unknown [15]. Though most of the earlier work assumes that the relay can transmit and listen over the same band, the half-duplex constraint (the relay cannot simultaneously transmit and receive at the same time in the same band) is taken into account in later work, for example [16, 17]. The information theoretic studies reveal that when there are one or two relays, the best strategy is to make use of both the source and relay transmissions at the user location, rather than multihopping from the source to

the destination through the relay(s). The intuition is that the user can make use of signals from both the source and the relay to get a better signal strength and hence a better rate. Multihopping on the other hand, ignores the signal from the source, however strong it is.

The information theoretic relaying protocols mentioned above often involve complicated multiuser coding and decoding techniques, that are too far from practical implementation. There have been some recent work trying to bridge the gap between the information theoretic and practical multihopping schemes, e.g., [18,19]. Most of the results in these works correspond to the case of a linear network of nodes, where there is a single commodity flow of message from the source node to sink node through a set of relay nodes. Any interference is only due to simultaneous transmissions from different relay nodes. This can be completely eliminated, by multiuser coding/decoding techniques. Such analysis does not carry over directly to the cellular systems since there are multiple simultaneous flows and multiuser techniques may incur significant overhead.

#### 4. **Fairness**

Service level agreements of cellular service providers entail certain fairness requirements. For example, in cellular system with voice users, the edge users and the users near the cell require the same level of service. Many other fairness schemes including proportional fairness [20] and max-min fairness [21, Chapter 6] have been proposed for cellular systems serving voice and data. Present day cellular systems implement schedulers in the MAC layer to provide various degrees of fairness to users. In this work, we assume that the 90% of users are required to be served at a common rate. When relays are present in the system, designing distributed scheduling schemes to provide fairness is an active area of research.

In this dissertation, we evaluate the performance of low-cost half-duplex relays in the downlink of a cellular system. The deployment scenario we consider is to mount a low-cost (preferably low-powered) device per sector over roof-tops of buildings. Such

devices can relay the information from the base station to users in the cell.

## 2.2 Related work and overview of our contribution

Relay deployment in a cellular system has been proposed to solve the issue of lack of coverage over a large area [22]. The use of relays in cellular systems have also been proposed to bring capacity improvements [23]. Viswanathan et al. [9] studied the performance of a centralized throughput-optimal scheduler on a cellular network with relays. They present a centralized downlink scheduling scheme that guarantees the stability of user queues for the largest set of arrival rates into the system. Each user has a queue at the base station and at its serving relay and the objective of the scheduler is to stabilize both queues while maximizing the throughput. The throughput results obtained by simulations in [9] suggest that simultaneous transmissions (due to multihopping) exploiting spatial reuse could lead to cell-wide throughput gains in a cellular network.

In addition to a multihopping model, wherein the message travels to the destination in two hops, in this dissertation we evaluate the performance of a *collaborative power addition* (CPA) scheme with a single relay available per user. We bring an additional dimension to the benefits of relays in a cellular system, by quantifying the power savings due to deployment of relays. Peak power savings in cellular networks are very important elements of amplifier costs in base stations. Significant peak power savings can reduce the cost of amplifiers and hence capital expenses for deploying cellular networks. Also, average power savings while operating cellular networks can save operational expenses such as electricity bills for the cellular operators.

In the first CPA scheme, we first consider a hypothetical model where 90% of the users are required to be served a file (henceforth, we use the term *message* and *file* interchangeably) of a fixed size within a certain deadline. Depending on the interference seen by each user, the mutual information (MI) or the instantaneous “rate” of the users vary over time. Users leave the system as they get the complete file within the deadline. We assume facility of large computation and use an offline computation to find the worst

10% of users that would not be able to get the complete file within the deadline and discard them at the beginning. We run the real system without the users in outage. When relays are present in the system, we evaluate the peak power savings at the base station to deliver a file of same size to the same number of users in the system as when the base stations and relays are transmitting at their peak power limits. We also find the improvements in common rate for the users in the presence of the relays when the peak power of the base stations are fixed for both the baseline and the system with relays.

We then include the power control capability to the base stations and relays. We evaluate the power savings and throughput improvement in the collaborative power addition scheme (PC-CPA). For a desired common rate requirement for 90% of users in the system, we find the common peak power constraint in the baseline case and in the system with relays to guarantee the common rate. When the relays get the complete message, they collaborate with the base station to transmit the message to the users. Each time a relay becomes eligible to transmit, the optimal set of powers are found to satisfy the desired rate requirement. The set of users that violate the peak power constraint are discarded at the beginning. The improvements in common rate are also evaluated through a similar procedure.

The second relaying scheme we evaluate is simple multihopping, wherein the base stations and relays transmit in *orthogonal* time slots. The baseline system is similar to the baseline in the CPA scheme described above. When relays are present in the system, we simulate a time-slotted system. For a common rate requirement for 90% of users in the system, the base stations transmit at peak power in odd time slots. The relays and users are in the receive mode. In the even time slots, the base stations are turned off and relays transmit to the respective users. The relays employ power control to target the remainder of the user population to provide the residual rate to the users. We describe the system in detail in the later sections.

We do not consider any multiuser scheduling gains, MIMO gains and any other complex interference mitigation techniques. Thus the gains shown in the network are purely due to the power gains at the user location due to the relay transmissions.

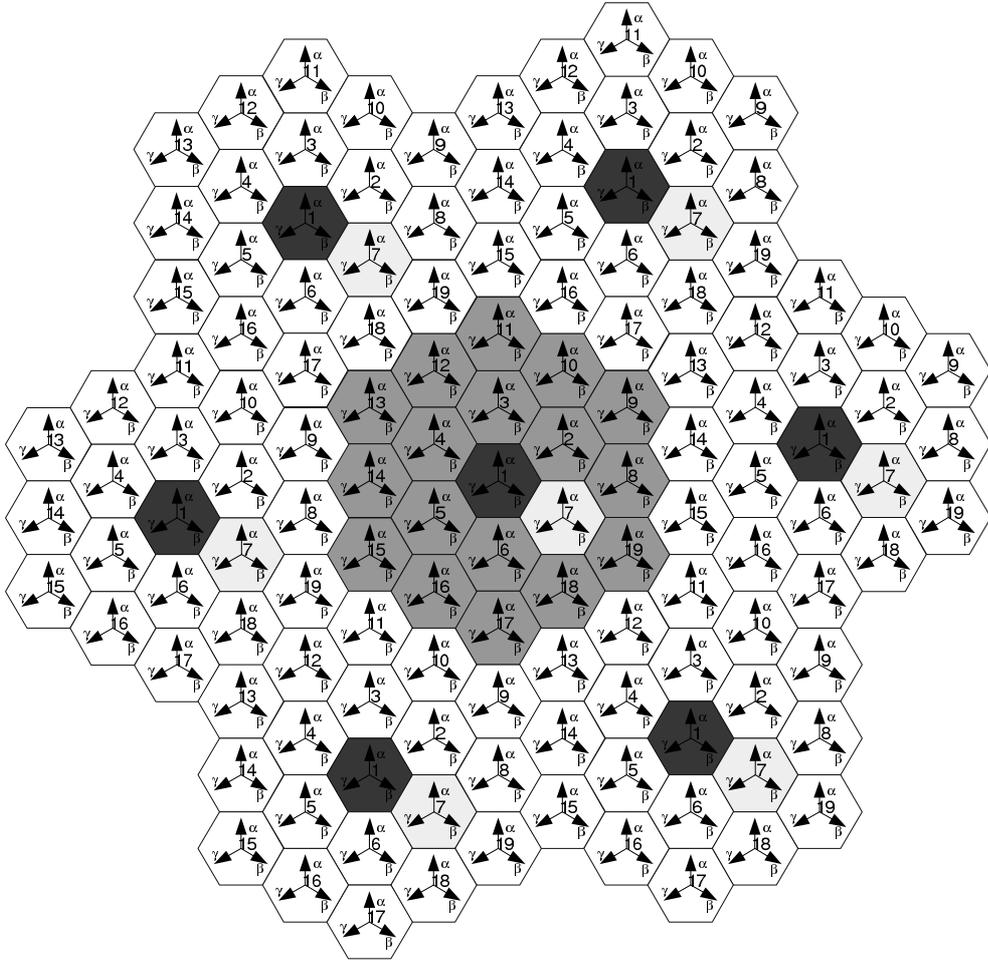


Figure 2.1: Wrap-around simulation model. The center ring of 19 cells are used for the simulation. The surrounding cell activity is mirrored in the center ring. The direction of the arrows represent the direction of the main lobe of the sectorized antenna.

### 2.3 The set-up

Our work aims to evaluate the power savings and improvement in common rate among users due to relay deployments in a cellular system. However, to model and simulate all dynamics of a cellular system can be too complicated. In order to overcome such difficulties, we make some reasonable simplifying assumptions and take an idealized look at the model and operation of a cellular system in our work. In order to make a fair comparison, the assumptions are kept consistent across systems with and without relays. We consider a cellular system with idealized hexagonal cells with a base station at the center of each cell. The topology is shown in Figure 2.1. The first two tiers of

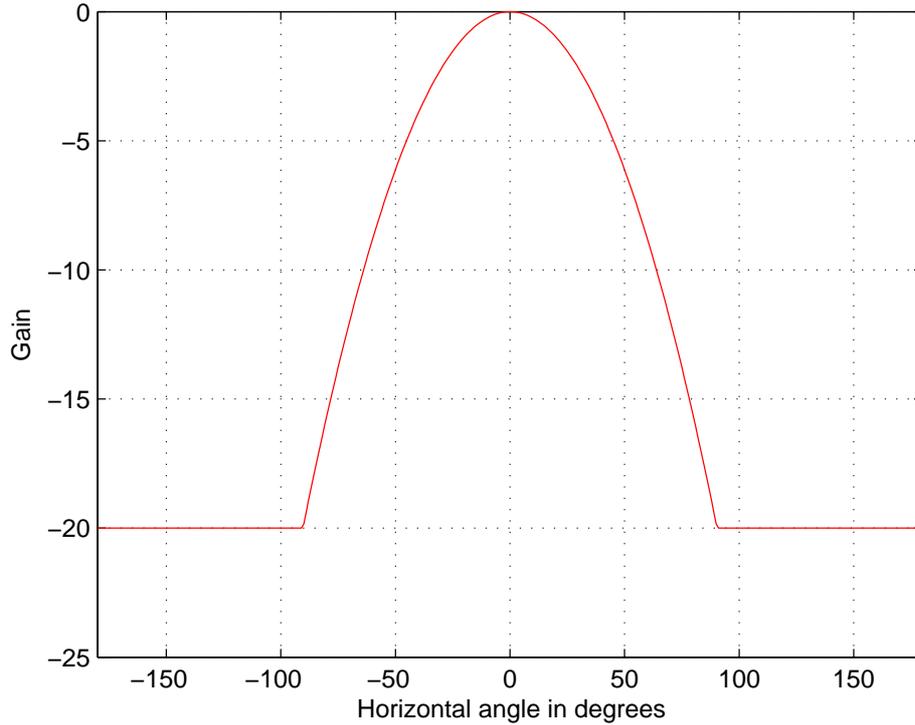


Figure 2.2: Antenna gain pattern (from [1]) as a function of the horizontal angle in degrees. The mathematical expression for the gain is given in equation (2.1).

interferers are considered and the activities of the farther tier of cells are mirrored by the center ring of 19 cells. The site-to-site distance (distance between any two base stations) is taken to be 1 mile. The cells are divided into 120 degree sectors, each sector illuminated by a base station antenna pattern given by

$$A(\theta) = -\min\left(12\left(\frac{\theta}{\theta_{3dB}}\right)^2, A_{\max}\right), \quad (2.1)$$

where  $A(\theta)$  is the antenna gain in dBi in the direction  $\theta$ ,  $-180 \leq \theta \leq 180$ ,  $\min(\cdot)$  denotes the minimum function,  $\theta_{3dB} = 70$  degrees is the 3 dB beamwidth and  $A_{\max} = 20$  dB is the maximum attenuation. The antenna gain pattern is shown in Figure 2.2.

At the receiving terminal (relay or user), the transmitted power undergoes attenuation due to the distance traveled and shadowing effects around the receiver. The propagation attenuation between a transmitting terminal (base station or relay) and a

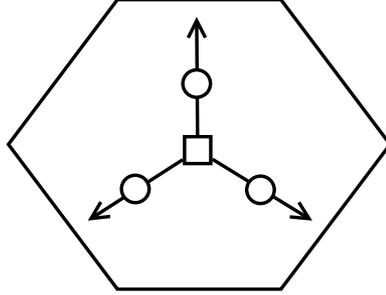


Figure 2.3: Position of relay location in a cell. The relays (represented by small circles) are placed at half the cell radius in the direction (given by the arrows) of the main lobe of the sector antenna. The base station at the center of the cell is represented by a square.

receiving terminal (relay or user) consists of the path loss and the shadowing component. At any receiving terminal, the transmitted power is attenuated in dB as  $PL(d) = -31.5 - 38 \log_{10} d$ , where  $d$  is in meters. The shadowing is modeled as lognormal with mean 0 dB and a standard deviation of 8 dB. The shadowing is assumed to be spatially uncorrelated and fixed for a given set of user locations. The base station and the relay antenna gains are taken to be 15 dB (at zero degree horizontal angle) and user antenna gain as  $-1$  dB. Other losses account for 10 dB. Together with the above losses, we include the antenna pattern loss to calculate the received power. The receiver noise figure is set at 5 dB, and the thermal noise power at the each receiving terminal (relay or user) is assumed to be  $-102$  dBm. The effect of multipath small scale fading is ignored in our simulations. All users share the same band of frequencies and hence simultaneous transmissions can interfere with each other. The total interference at each receiving terminal from all transmitters in the system is modeled as Gaussian noise and idealizes that other users use Gaussian codebooks. The achievable rate to a user  $i$  at time  $t$  is calculated as the Shannon rate

$$R_i(t) = \log_2(1 + \rho_i(t)), \quad (2.2)$$

where  $\rho_i(t)$  denotes the SINR for user  $i$  at time  $t$ . The parameters used in the above mentioned simulation set-up are summarized in Table 2.1. We use this simulation set-up

Table 2.1: Parameters used in simulations

Network Topology	19 cells, 3 sectors per cell with wraparound
Site-to-site distance	1 mile
Bandwidth	5 MHz
Path loss model	COST-231 Hata model
Path loss exponent	$\alpha = 3.8$
Shadowing	Lognormal, with zero mean, 8 dB standard deviation for access and backhaul
Multipath fading	None
Antenna Pattern	Sectorized for base stations Omnidirectional for relays
Antenna gains	15 dB (for base station and relays) -1 dB for users
Other losses	10 dB
Thermal noise power at the receiver	-102 dBm
Outage	10% for baseline and with relays

for evaluating all relaying methodologies proposed in this dissertation.

### 2.3.1 Placement of users and relays in the system

We simulate a downlink OFDM-like system wherein users in orthogonal time or frequency slots do not interfere with each other. However, users in the same resource unit interfere with the other transmissions in the band. We simulate the worst case scenario where the system is fully loaded, i.e., users are present in all available resource units (or time-frequency slots) in all the sectors. The time-frequency slots are reused in each sector. We assume that the time-frequency slots are orthogonal, and focus only on a particular time-frequency slot within which we simulate the complete

cellular system such that there is one active user per sector at a given time. Hence, in a 19-cell network with 3 sectors per base antenna, at most 57 users are served at a given time-frequency slot. In our simulations, we use the following heuristic to create a random user population along with an association rule. Users are placed one-by-one in a uniformly random fashion across the network until all 57 base station sectors are occupied. For each random realization of a user location, the base station sector with the highest received signal strength is chosen to associate with the user. If the base station sector is already occupied by another user, the user is not allowed into the system and a new user location is generated. Along with a random realization of a user location, independent lognormal random variables also instantiated to account for the shadow fading gains between each base station and the user in the baseline system. If relays are present in the system, the fading gains are also generated for base station – relay links and relay – relay links. In this way, the random placement is carried out until all 57 sectors are occupied by exactly one user per sector. Each user is equipped with an omni-directional antenna.

A relay with an omni-directional antenna is placed in the direction of the main lobe of each base station sector antenna as shown in Figure 2.3. The relays always associate with the corresponding base station sector. The relay placement is an important parameter to be considered since the power gains and throughput improvements depend on the interference generated by the relays, which in turn, depends on the transmit power, geographic location of the relays and the propagation environment. In our simulations, we experiment with various relay placements and the simulation results are presented for the relay locations for which the gains are found to be maximum. The relay powers are also varied so that we get the maximum peak power savings.

## 2.4 Organization

The organization of Section numbers for half-duplex relaying methodologies and the performance evaluation is given in the Table 2.2.

Table 2.2: PERFORMANCE EVALUATION OF RELAYS IN DOWNLINK CELLULAR SYSTEMS

(2.5) Collaborative Power Addition (CPA)		(2.8) Orthogonal Relaying
(2.6) CPA with Peak power transmissions (P-CPA)	(2.7) CPA with power control (PC-CPA)	
(2.6.1) Principle of operation – Baseline – Relay	(2.7.1) Principle of operation – Baseline – Relay (2.7.2) Optimization Framework – min average power – min peak power – rate improvement	(2.8.1) Network operation – Baseline – Relay (2.8.2) User discarding method – Baseline – Relay
(2.6.2) User discarding method	(2.7.3) User discarding method	
(2.6.3) Network operation	(2.7.4) Network operation	
(2.6.4) Simulation Results	(2.7.5) Simulation Results	(2.8.3) Simulation Results
– Power savings	– Power savings	– Power savings
– Rate gains	– Rate gains	– Rate gains

## 2.5 Collaborative relaying in cellular networks (CPA)

In the Collaborative Power Addition (CPA) scheme devised in [24], the relay collaborates with the base station to help the message reach to the destination. In our simulation model, each base station sector has a single user to be served and a relay that may help the source to deliver the message to the user associated with the source. In what follows, we focus our attention on an isolated triplet of base station (source), relay and user in a single sector. Gaussian encoding is used across all other sectors, the interference from other sectors is considered as if it were additive Gaussian noise. Suppose the source wants to transmit one of  $M$  messages to the destination, under a power constraint  $P$ . The source transmits a Gaussian codeword of length  $N = (\log M)/R$ , where  $R$  is the rate of the code. By Shannon's channel coding theorem [25, Chapter 9], if  $N$  is large enough, the message can be decoded reliably at the destination provided  $R < \log(1 + \rho)$ , where  $\rho$  is the received SINR. In our simulations, we are interested in achievable rates and assume that the instantaneous mutual information at the receiver is exactly  $R = \log(1 + \rho)$ .

Assume that the source picks a rate  $R$  code  $\mathcal{C}_1$  and sends one of  $M$  equally probable messages to the destination, using a codeword of length  $N$ . Let the received SINR  $\rho_{SR}$  between the source and relay be greater than the received SINR  $\rho_{SD}$  at the destination. Then, there exists some  $\beta > 1$  such that

$$\log(1 + \rho_{SR}) = \beta \log(1 + \rho_{SD}), \quad (2.3)$$

i.e., the capacity of the channel from source to relay is  $\beta$  times greater than the channel from source to destination. We can now construct codebook  $\mathcal{C}_2$  derived from  $\mathcal{C}_1$  by observing only the first  $\lceil N/\beta \rceil$  symbols of every codeword. The relay can then reliably decode the received message since the rate of  $\mathcal{C}_2$  is

$$R' = \frac{\log M}{\lceil N/\beta \rceil} < \log(1 + \rho_{SD}) \quad (2.4)$$

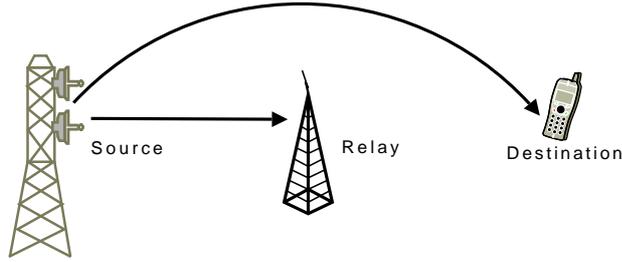


Figure 2.4: Collaborative relaying: before relay decodes the message

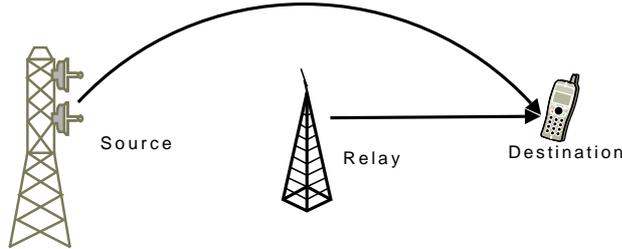


Figure 2.5: Collaborative relaying: after relay decodes the message

In [26, Appendix F], the authors discuss the coding interpretation of a similar collaborative strategy. The authors also discuss the connection of such a coding setting with coding for arbitrary varying channel (AVC), which was first dealt with in [27] and then subsequently studied in [28]. We simulate a similar collaborative coding strategy wherein before the relay decodes the message as shown in Figure 2.4, the received power at the destination node is only due to the base station transmission. After the relay decodes the message, the relay joins the base station to help the base station in delivering the message to the destination as shown in Figure 2.5. At this point, if we assume that transmit symbol time slots at the relay and base station are synchronized and the code books are shared, the system can be viewed as a  $2 \times 1$  MISO (Multiple-Input Single-Output) system without channel information at the transmitter. There is an effective power addition of the base station and relay transmissions at the destination [29, Chapter 3]. A similar scheme was proposed in the literature dynamic decode and forward (DDF) scheme [30].

We simulate this collaborative relaying strategy in two ways:

- Base station and relay transmit at their respective peak powers. In this case, the transmit power is fixed and the users get variable rates depending on SINR at

the user locations. When a target rate is obtained by a user, the user leaves the system and the corresponding base station sector is turned off, thus reducing the amount of interference in the system. We term this the peak collaborative power addition (P-CPA) scheme. This is described in Section 2.6.

- Base station and relay operate with power control so that the users obtain a target desired rate. In the baseline case, for a given desired rate requirement  $r_0$  bps/Hz, a feasible set of powers are found to better satisfy the rate requirement, allowing for a certain users to be in outage. When the relays decode the message in the collaborative scheme, the optimal powers are recalculated to find another feasible set of powers to satisfy the rate requirement at the same outage level. We term this the power control collaborative power addition (PC-CPA) scheme. This is described in Section 2.7.

## 2.6 CPA with peak power transmissions (P-CPA)

### 2.6.1 Principle of operation

#### P-CPA Baseline

In the baseline of the P-CPA scheme, each base station sector transmits at its peak power to its own intended user. Since all users share the same band of frequencies, they observe interference from all the base station sectors in the system. If at time  $t$ ,  $p_i(t)$  is the peak power of the transmitting base station sector corresponding to the  $i$ th user and  $h_{ij}$  is the channel gain, including path loss and shadowing, from the  $j$ th base to the  $i$ th user and  $\sigma^2$  is the variance of the noise power at the receiver, the instantaneous received SINR for user  $i$  is given by

$$\rho_i(t) = \frac{h_{ii}p_i(t)}{\sum_{j \neq i} h_{ij}p_j(t) + \sigma^2}. \quad (2.5)$$

Since we assume Gaussian signaling, the MI (mutual information) or the instantaneous “rate” to each user is given as

$$R_i(t) = \log_2(1 + \rho_i(t)) \text{ bits/symbol.} \quad (2.6)$$

At time  $t = 0$ , all base stations simultaneously transmit to their associated user. As time progresses, for any given time interval  $[t, t + \Delta t]$ , user  $i$  accumulates MI  $I_i(\Delta t) = R_i(t)\Delta t$ . The MI for user  $i$  at time  $t$  is given by,

$$I_i(t) = \int_0^t \log_2(1 + \rho_i(\xi)) d\xi. \quad (2.7)$$

If user  $i$  accumulates MI corresponding to the required amount  $L$  of data before the deadline  $T$ , i.e.,

$$\tau_i = \min_{0 \leq t \leq T} \{t : I_i(t) = L\}. \quad (2.8)$$

then the user leaves the system and his associated base station sector is turned off at time  $\tau_i$ , reducing the overall interference levels in the system. Hence,

$$p_i(t) = \begin{cases} P, & t < \min(\tau_i, T), \\ 0, & t \geq \min(\tau_i, T), \end{cases} \quad (2.9)$$

where,  $P$  is the peak power of the base station transmission. Note that the  $\rho_i(t)$  of user  $i$  and the rate  $R_i(t)$  are time varying quantities. At time  $t = T$ , the users that remain in the system are those users that did not get the complete file. It is these remaining users that are ascribed to be in *outage*.

### **P-CPA system with Relays**

The operation of the P-CPA system with relays is as follows. The requirement is the same as the baseline case: to deliver a file of size  $L$  to as many users within the time  $T$ . At time  $t = 0$ , the base stations transmit at peak power to users associated with them. The relay node placed in the sector also receives the data sent to the user by the base station. If the relay gets the complete file before the user gets it, the relay

can potentially be useful to the user by helping it get the message faster. On the flip side, the relay transmission can create additional interference for the other users in the system. In our simulations, we follow a myopic<sup>1</sup> policy on whether to turn on the relay or not: the relay transmits at peak power to help its user only if the instantaneous sum-rate of the whole system increases by turning the relay on. The sum-rate of the system is calculated as the sum total of the instantaneous rates of the existing users in the system and is a natural system-wide metric to use in order to decide whether the relays should transmit or not. At every epoch, a relay gets the message, among the set of all relays that are eligible to be turned on, the myopic sum-rate metric is applied and those relays that increase the sum-rate are turned on to help the users in the system.

If the relay increases the sum-rate of the system, the relay is turned on and helps the user with a transmission reinforcing the same message as the base station using the code described in Section 2.5. If  $q_i(t)$  is the power transmitted from the relay  $i$  at time  $t$  and  $g_{ij}$  is the channel gain from the user  $i$  to the relay  $j$ , the effective SINR at  $i^{\text{th}}$  user location when the relay is active is given by

$$\rho_i^{\text{relay}}(t) = \frac{h_{ii}p_i(t) + g_{ii}q_i(t)}{\sum_{j \neq i} h_{ij}p_j(t) + g_{ij}q_j(t) + \sigma^2}. \quad (2.10)$$

The instantaneous rate and the mutual information for user  $i$  at time  $t$  are given by

$$R_i^{\text{relay}}(t) = \log_2(1 + \rho_i^{\text{relay}}(t)) \quad (2.11)$$

$$I_i^{\text{relay}}(t) = \int_0^t R_i^{\text{relay}}(\xi) d\xi. \quad (2.12)$$

If  $H_{ij}$  denote the channel gain from the  $j^{\text{th}}$  base station to the  $i^{\text{th}}$  relay,

$$J_i(t) = \int_0^t \log_2 \left( 1 + \frac{H_{ii}p_i(\xi)}{\sum_{j \neq i} H_{ij}p_j(\xi) + \sigma^2} \right) d\xi \quad (2.13)$$

represents the cumulative MI at the relay at time  $t$ .

Suppose the relay  $i$  becomes eligible to transmit at time  $t$ , i.e.,  $J_i(t) > L$ , then

---

<sup>1</sup>The policy is myopic since, at the time when the relay gets the message, the global optimal decision whether the relay should transmit or not is unknown.

denote the sum-rate of the system at time  $t$  as a function of  $q_i(t)$  as

$$SR(t, q_i(t)) = \sum_i \log_2 \left( 1 + \frac{h_{ii}p_i(t) + g_{ii}q_i(t)}{\sum_{j \neq i} h_{ij}p_j(t) + g_{ij}q_j(t) + \sigma^2} \right). \quad (2.14)$$

Then, the relay power at time  $t$  is given by

$$q_i(t) = \begin{cases} Q, & \text{if } J_i(t) > L, SR(t, Q) > SR(t, 0) \text{ and } t < T, \\ 0, & \text{otherwise,} \end{cases} \quad (2.15)$$

where  $Q$  is the peak power constraint of the relays. Each user sees a time-varying SINR and the time-varying rate given by  $R_i(t) = \log_2(1 + \rho_i^{relay}(t))$ . As with the baseline case, for any interval of time  $[t, t + \Delta t]$ , user  $i$  accumulates MI amounting to  $I_i(\Delta t) = R_i(t)\Delta t$  and the MI for user  $i$  at time  $t$  is

$$I_i(t) = \int_0^t \log_2(1 + \rho_i^{relay}(\xi)) d\xi. \quad (2.16)$$

Similar to the baseline case, if the user accumulates MI amounting to the full file size  $L$  within the stipulated time  $T$ , the user leaves the system and the associated base station and relay are switched off. Thus the effective interference in the system is reduced. At time  $t = T$ , the users that remain in the system are those users that did not get the complete file.

### 2.6.2 User discarding methodology

The user discarding procedure can be divided in two phases:

- (i) A *learning phase* where we learn the power threshold, that is used as a criterion to determine the users in outage. All the users that are not in outage require their corresponding base stations to have peak powers lower than the power threshold. The network will be operated with peak powers of all base stations are capped at the power threshold.
- (ii) After the power threshold is found, we assume the availability of very fast computing facility and perform an off-line computation to find out the set of users

that are in outage. Such users could be discarded upfront before the start of the simulations so that the others users benefit from the absence of interference from these users. Hence, this takes care of the causality of the discarding phase from the operation of the real network.

We conduct Monte Carlo simulation runs for the baseline case as well as the case with relays in our work. For each simulation run random instantiations of 57 user locations (as per Section 2.3.1) and associated statistically independent shadow fading values are generated. Once the random values are instantiated, these values are stored in our simulation software program. The same set of user locations and shadow fading values serve as inputs to the baseline and the system with relays.

We explain the learning phase now. Consider a single instance of the simulation runs in the baseline case. For the given instantiation, there are 57 users one in each sector. We fix a peak power threshold  $P$  for the base stations and also fix the desired common rate for users as  $r_0$  bits/sec/Hz. When the baseline system operation is over, the users that are in outage remain in the system at time  $T$ . Let the number of users in outage for the  $k^{th}$  instantiation when the power threshold is  $P$  and desired common rate  $r_0$  be  $O_k(P, r_0)$ . For the same power threshold  $P$  and common rate  $r_0$ , we run a large number  $K$  of instantiations. We then find the total number of users in outage as

$$O(P, r_0) = \sum_{k=1}^K O_k(P, r_0). \quad (2.17)$$

The percentage of users in outage for the threshold  $P$  and desired rate  $r_0$  is then

$$\frac{O(P, r_0)}{57K} \times 100 \% \quad (2.18)$$

If  $O(P, r_0) > 10\%$ , we increase the power threshold to  $P' > P$ . On the other hand, if  $O(P, r_0) < 10\%$ , we decrease the power threshold to  $P'' < P$ . Proceeding in this fashion, the base station peak power thresholds are adjusted such that exactly 90% of the users are guaranteed the desired rate of  $r_0$  bits/sec/Hz and the rest of the 10% users are in outage.

We could improve the performance of the system by discarding the users in outage upfront, since the interference due to the presence of these users will be eliminated at time  $t = 0$ . In our simulations, for a large user population over  $K$  instantiations, we identify 10% users in outage<sup>2</sup> by first running  $K$  instantiations of the system with all the users present in the system. We store the coordinates of all the users that were in outage at the end of each of the  $K$  instantiations. We then eliminate the outage users from the system (by preserving the coordinates of the user locations of only those users not in outage for all the  $K$  instantiations) at time  $t = 0$  in the real network simulations. Thus, the existing users in the system would experience lesser interference due to the absence of those users in outage when the real network is simulated.

### 2.6.3 Network Operation and Simulation Aspects

Our objective is to obtain power savings and throughput improvement benefits due to deployment of relays in cellular system. To compare systems with and without relays in the CPA based relaying scheme, we simplify the operation of a cellular downlink system such that 90% of the users in the system are guaranteed to be delivered a file of fixed size  $L$ , within a fixed period of time  $T$ . The file could be different for all users but the file sizes are fixed. Such an operation brings in the notion of a common rate for the users in the system. In order that the system benefits from the users that get the message within the fixed time  $T$ , the satisfied users leave the system, thus no longer causing interference to the remaining users. The remaining 10% of the users that are not guaranteed of the file of size  $L$  are ascribed to be in *outage*.

In our simulations, for the sake of simplicity, all base stations are assumed to have the same peak power threshold values. We run  $K = 200$  (amounting to 11400 user instantiations) different user instantiations in the system. The common rate requirement is set as 1 bit/sec/Hz. This common rate requirement translate to 0 dB common target SINR requirement. We divide the total time  $T$  into 1000 mini-slots and at the end of

---

<sup>2</sup>We remark that the eliminated set of 10% users in outage is not claimed to be the optimum set as would be obtained by evaluating all possible subsets amounting to 10% of the users. The latter is computationally prohibitive.

each mini-slot, we keep track of the cumulative MI  $I_i(t)$  of each user  $i$ . If at the end of a mini-slot, a particular user's cumulative MI exceeds the file size  $L$ , the base station corresponding to that user is turned off.

We run the baseline for different peak power values of the base station (5 W to 30 W in increments of 5 W). For each peak power value, the relay powers are varied as a factor of the base station power. Figure 2.6 shows the variation of outage probability for various base station powers and various relay powers. For the case when there are no relays in the system (ratio of relay power to base station power is zero), increasing the peak powers of the base station decreases the outage. The percentage of outage saturates below a certain threshold as the interference limit sets in. As we increase the relay powers by increasing the ratio of relay power to base station power, the outage reduces but quickly saturates to a certain threshold outage value, because of the interference limit. From the Figure 2.6 it is clear that interference limit is quickly reached and limits the performance of system with relays. This is because we do not control the interference and peak power transmissions from the base stations and relays lead to a highly interference limited scenario.

#### 2.6.4 Simulation results

##### Power savings

The base station peak power required to guarantee 90% of the users (after the 10% users in outage have been removed) a rate of 1 bit/sec/Hz is 21 W in the baseline case and it requires 15 W for a system with relays. The relays transmit 1 W of peak power. Hence the peak power savings at the base station locations in this case is 1.46 dB as shown in Table 2.3.

##### Rate gains

In order to evaluate the throughput improvement, we find how much the common rate of 90% of users can be improved with the peak power of the base stations being fixed. For the baseline, we fix the power of the base stations to 21 W, so that 90% of the

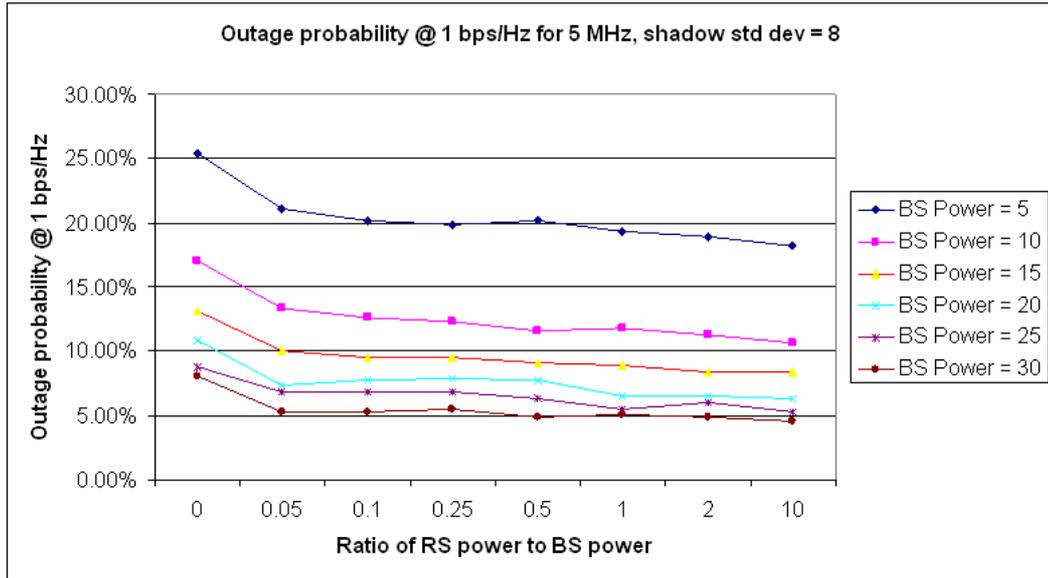


Figure 2.6: Variation of outage with relay powers and base station powers for the P-CPA scheme. As we increase the base station powers with no relays in the system (ratio = 0), the outage decreases and saturates at around 5%, due to the interference limit. The interference limit sets in very quickly even for smaller values of relay powers.

users are guaranteed to get 1 bit/sec/Hz (as obtained in the previous section). The 10% of the users in outage are eliminated as explained in Section 2.6.2. For the P-CPA system with relays, the peak power threshold of the base stations are fixed to 21 W (the same value as in the base line case). For the same peak power for the base stations and with relays present in the system, we expect the common rate to be better than 1 bits/sec/Hz. To find the improvement in common rate, we fix a desired common rate  $r' > 1$  bit/sec/Hz and run the system with relays. If this desired common rate is feasible<sup>3</sup>, we double the desired common rate and run the simulations again. Else, if the desired common rate is infeasible, we fix the new desired common rate at half the difference between the highest feasible common rate and the lowest infeasible common rate and rerun the simulations. In this manner, we converge to the achievable common rate in the presence of relays. In our simulations, we find that the common rate can be improved to 1.21 bits/sec/Hz in the CPA based relaying scheme. Hence the common

<sup>3</sup>The common rate is feasible if all the users present in the system are able to get the desired rate.

Table 2.3: P-CPA relaying. Base station and relays transmit at peak power

Peak power required to guarantee 1 bps/Hz at 10% outage Baseline (No relays)	Peak power required to guarantee 1 bps/Hz at 10% outage With relays	Savings in dB
21 W	15 W	1.46
Common rate for 90% users Baseline (No relays)	Common rate for 90% users With relays	Percentage rate increase
1 bps/Hz	1.21 bps/Hz	21 %

rate improvement is 21%.

## 2.7 Power control based collaborative relaying (PC-CPA)

In case of P-CPA relaying in Section 2.6.3, we observed that the interference from the other relays and base station sectors was limiting the peak power savings in the system with relays. The reason for that is when the relays transmit to help the users, they transmit with peak powers and hence increase the interference levels in the system. If we could find the optimal set of powers to transmit for the base station and the relays, we could reduce the overall interference levels in the system. This may improve the gains in the system.

In the following, we describe a framework for power control in the downlink of a cellular system with relays. When the relays are not present, downlink power control in a cellular system is well studied and understood [31]. When relays are present in the system, power control, if performed jointly at the base stations and relay locations, can provide power savings and throughput improvement. We describe the PC-CPA relaying scheme in the following sections.

### 2.7.1 Principle of operation

#### PC-CPA Baseline

For the PC-CPA baseline, the aim is to deliver a desired common rate for 90% of the user population by employing a simple power control scheme. Each base station sector powers down its transmitted power within the peak power limitations so that 90% of users are guaranteed a desired common rate of  $r_0$  bits/sec/Hz. Since all users share the same band of frequencies, they observe interference from all the base station sectors in the system. We use a common subscript for a base station or a user in a particular sector. If at time  $t$ ,  $p_i$  is the power of the transmitting base antenna corresponding to the  $i$ th user (we drop the argument  $t$ ) and  $h_{ij}$  is the channel gain, including path loss and shadowing, from the  $j$ th base to the  $i$ th user and  $\sigma^2$  is the noise power at the receiver, the instantaneous SINR of the  $i$ th user in the system is given by

$$\rho_i(t) = \frac{h_{ii}p_i}{\sum_{j \neq i} h_{ij}p_j + \sigma^2}. \quad (2.19)$$

Since the transmission use Gaussian codebooks, the corresponding instantaneous rate for the user  $i$  is given by

$$R_i(t) = \log_2(1 + \rho_i(t)) \text{ bits/sec/Hz}. \quad (2.20)$$

The set of feasible powers such that the users not in outage are guaranteed with a rate  $r_0$  is obtained by solving for the feasibility of instantaneous rates subject to peak power constraints, specified by

$$\log_2(1 + \rho_i(t)) \geq r_0, \quad (2.21)$$

$$\text{i.e., } \rho_i(t) \geq 2^{r_0} - 1, \quad (2.22)$$

$$\text{subject to } p_i \leq p_{i,\max}, \quad (2.23)$$

for all users  $i$  not in outage. In practice, each base station increases its power autonomously in small increments, until it hits the peak power limit or when the user

associated with it attains the desired rate  $r_0$ . Users that make the power constraint to go active before attaining the desired rate are discarded. We simulate the system without the users in outage such that all users get the desired rate. Since the transmit powers of the base stations are such that all users get a common rate, none of the users leave the system.

### **PC-CPA System with Relays**

In the PC-CPA system with relays, 10% of users are discarded in the manner similar to the baseline system. At time  $t = 0$ , the relays do not have the complete message required to relay to the user. Hence, the system starts out as it does for the baseline case. The base stations increase their powers autonomously in small increments targetting the users rates to increase. Users that do not meet the peak power constraints in (2.23) are eliminated one after the other. The remaining users get the desired rate without violating the peak power constraint at the base stations. While the base station transmissions are targetted to the users, the relay in each sector also listens to the transmission by the base stations. Depending on the channel conditions and coupling of interference from the adjacent sectors, the relays get their message at different points in time. When the relay in the sector decodes the message from the base station, the relay collaboratively helps the base station such that the user gets a rate corresponding to the total SINR from the relay and the base station. As described in Section 2.5, the code books at the base stations and relays are designed such that the mutual information at the receiver corresponds to the sum of received powers at the user location [24]. In order to maintain the common desired rate for all users, the relay and base station jointly adjust their powers so that the user gets the desired rate. This ensures that the base station and relay transmit just enough power to the user to obtain the desired rate.

Let  $p_i$  denote the power transmitted by the base station sector  $i$  at time  $t$  and  $h_{ij}$  be the channel gain, including path loss and shadowing, from the  $j$ th base to the  $i$ th user. Let  $q_i$  be the power transmitted from the relay  $i$  at time  $t$  and  $g_{ij}$  be the channel gain to the user  $i$  from the relay  $j$ . Then, when the relay and base station transmit

simultaneously, the effective SINR at  $i^{th}$  user location when the relay is active is given by

$$\rho_i^{relay}(t) = \frac{h_{ii}p_i + g_{ii}q_i}{\sum_{j \neq i} h_{ij}p_j + g_{ij}q_j + \sigma^2}, \quad (2.24)$$

As with the baseline case, the set of feasible powers (for both base station antennas and relays) such that the users not in outage are guaranteed with a rate  $r_0$  is obtained by solving for the feasibility of instantaneous rates subject to peak power constraints

$$\log_2(1 + \rho_i^{relay}(t)) \geq r_0, \quad (2.25)$$

$$\text{i.e., } \rho_i^{relay}(t) \geq 2^{r_0} - 1, \quad (2.26)$$

$$\text{subject to } p_i \leq p_{i,\max}, \quad (2.27)$$

$$\text{and } q_i \leq q_{i,\max}, \quad (2.28)$$

for all users  $i$  not in outage. If we consider the transmit powers of the base stations and relays as variables of optimization, we have a total of  $2N$  variables, for  $N$  base station sectors in the system. Thus, power control in cellular systems in the presence of relays gives us additional  $N$  degrees of freedom to optimize over. The transmit powers in the system can be optimized to reduce the maximum peak power transmission in the system, reduce total energy in the system etc. In what follows, we assume that a central controller has the knowledge of the all the channel gains between the base stations as well as relays and the users. We explain ways to achieve various aforementioned objectives using linear program (LP) formulations.

## 2.7.2 Optimization framework

### Minimizing the total instantaneous transmit powers

We are interested in evaluating the benefits of relays in minimizing the total instantaneous sum power in the system while delivering the common rate  $r_0$  with 10% of the users being omitted from the system. The practical benefit of minimizing the total sum of transmit powers in a cellular system is to save the energy costs in the network. Saving energy costs translate to saving electricity bills at the cell sites for the cellular

service provider.

The desired common rate for the users is fixed at  $r_0$  bits/symbol. We define  $\mathcal{A}(t)$  as the set of all active relays at time  $t$ , i.e., the set of relays that have obtained the message and are ready to help the base station.  $\mathcal{A}^c(t)$  denotes the complementary set of all inactive relays. For simplicity, we drop the argument and write  $p_i$  and  $q_i$  for the base station powers  $p_i(t)$  and  $q_i(t)$  at time  $t$ , respectively.

At a given time  $t$ , we solve the following optimization problem:

$$\min_{\substack{p_1, \dots, p_N \\ q_1, \dots, q_N}} \sum_i p_i + q_i \quad (2.29a)$$

$$\text{subject to} \quad \log_2 \left( 1 + \frac{h_{ii}p_i + g_{ii}q_i}{\sum_{j \neq i} h_{ij}p_j + g_{ij}q_j + \sigma^2} \right) \geq r_0, \quad i = 1, \dots, N, \quad (2.29b)$$

$$0 \leq p_i \leq p_{i,\max}, \quad i = 1, \dots, N, \quad (2.29c)$$

$$0 \leq q_i \leq q_{i,\max}, \quad i \in \mathcal{A}(t), \quad (2.29d)$$

$$q_i = 0, \quad i \in \mathcal{A}^c(t). \quad (2.29e)$$

The solution to the optimization problem (2.29),  $p_i^*, q_i^*, i = 1, \dots, N$  defines the powers  $p_i(t) = p_i^*$  and  $q_i(t) = q_i^*$  that are used at time  $t$ . The optimization problem (2.29) is an LP, since we can write the constraint (2.29b) as

$$\frac{1}{2^{r_0} - 1} (h_{ii}p_i + g_{ii}q_i) - \sum_{j \neq i} (h_{ij}p_j + g_{ij}q_j) \geq \sigma^2, \quad (2.30)$$

for  $i = 1, \dots, N$ . Rewriting (2.29) in vector form, we have

$$s^*(t) = \min_{\mathbf{p}, \mathbf{q}} \quad \mathbf{1}^T (\mathbf{p} + \mathbf{q}) \quad (2.31a)$$

$$\text{subject to} \quad \mathbf{A}\mathbf{p} + \mathbf{B}\mathbf{q} \leq -\sigma^2 \mathbf{1} \quad (2.31b)$$

$$\mathbf{0} \leq \mathbf{p} \leq \mathbf{p}_{\max}, \quad (2.31c)$$

$$\mathbf{0} \leq \mathbf{q} \leq \mathbf{q}_{\max}, \quad (2.31d)$$

where,

$$\mathbf{A} = \begin{pmatrix} -h_{11}/(2^{r_0} - 1) & h_{12} & \cdots & h_{1N} \\ h_{21} & -h_{22}/(2^{r_0} - 1) & \cdots & h_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ h_{N1} & h_{N2} & \cdots & -h_{NN}/(2^{r_0} - 1) \end{pmatrix}, \quad (2.32)$$

$$\mathbf{B} = \begin{pmatrix} -g_{11}/(2^{r_0} - 1) & g_{12} & \cdots & g_{1N} \\ g_{21} & -g_{22}/(2^{r_0} - 1) & \cdots & g_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ g_{N1} & g_{N2} & \cdots & -g_{NN}/(2^{r_0} - 1) \end{pmatrix}, \quad (2.33)$$

and

$$\begin{aligned} \mathbf{p}(t) &= [p_1(t) \cdots p_N(t)]^T, \\ \mathbf{q}(t) &= [q_1(t) \cdots q_N(t)]^T, \\ \mathbf{p}_{\max} &= [p_{1,\max} \cdots p_{N,\max}]^T, \\ \mathbf{q}_{\max} &= [q_{1,\max} \cdots q_{N,\max}]^T, \end{aligned} \quad (2.34)$$

where,  $q_{i,\max} = 0$ ,  $i \in \mathcal{A}^c(t)$ . Solution to the above LP provides the optimal power values that minimize the instantaneous total power in the system. We take a myopic approach of minimizing the total sum power of the system at time  $t$  in order to reduce the total average power transmission in the system. Each time a relay becomes eligible for transmission, the LP is solved to find the best set of powers by minimizing the instantaneous powers in the system. Note that in some cases, when a relay is eligible to help the base station, turning off the base station may be the optimal thing to do. This choice comes out as a solution to the optimization program.

### Minimizing the peak transmit power

Minimizing the peak transmit power leads to peak power savings in the system. A practical benefit of peak power savings is the significant savings in the cost of power amplifiers for the cellular base stations. If by deploying low-power relays in the system, we save in the cost of the power amplifiers of the base stations, cellular operators could save in capital expenses. To this end, we solve the following optimization problem of minimizing the maximum instantaneous transmit powers at the base stations:

$$\min_{\substack{p_1, \dots, p_N \\ q_1, \dots, q_N}} \max_i p_i \quad (2.35a)$$

$$\text{subject to} \quad \log_2 \left( 1 + \frac{h_{ii}p_i + g_{ii}q_i}{\sum_{j \neq i} h_{ij}p_j + g_{ij}q_j + \sigma^2} \right) \geq r_0, i = 1, \dots, N, \quad (2.35b)$$

$$0 \leq p_i \leq p_{i,\max}, \quad i = 1, \dots, N, \quad (2.35c)$$

$$0 \leq q_i \leq q_{i,\max}, \quad i \in \mathcal{A}(t), \quad (2.35d)$$

$$q_i = 0, \quad i \in \mathcal{A}^c(t). \quad (2.35e)$$

Rewriting the above LP in vector form yields for any dummy variable  $\alpha$ ,

$$p^*(t) = \min_{\mathbf{p}, \mathbf{q}} \alpha \quad (2.36a)$$

$$\text{subject to} \quad \mathbf{A}\mathbf{p} + \mathbf{B}\mathbf{q} \leq -\sigma^2\mathbf{1} \quad (2.36b)$$

$$\mathbf{0} \leq \mathbf{p} \leq \mathbf{p}_{\max}, \quad (2.36c)$$

$$\mathbf{0} \leq \mathbf{q} \leq \mathbf{q}_{\max}, \quad (2.36d)$$

$$\alpha\mathbf{1} \geq \mathbf{p}_{\max}, \quad (2.36e)$$

where,  $\mathbf{A}$  and  $\mathbf{B}$  are given by (2.32) and (2.33) respectively.

### Improving the common rate

As a corollary to the above approaches, if we keep the peak power constant across both the baseline and the system with relays, we can increase the common targeted rate in

the system with relays. The problem of maximizing the common rate can be posed as an optimization program with the transmit powers of the base station and relays as the variables. The “genie” then solves the optimization program:

$$\max_{p_1, \dots, p_N} r_0 \quad (2.37a)$$

$$\text{subject to} \quad \log_2 \left( 1 + \frac{h_{ii}p_i + g_{ii}q_i}{\sum_{j \neq i} h_{ij}p_j + g_{ij}q_j + \sigma^2} \right) \geq r_0, i = 1, \dots, N, \quad (2.37b)$$

$$0 \leq p_i \leq p_{i,\max}, \quad i = 1, \dots, N, \quad (2.37c)$$

$$0 \leq q_i \leq q_{i,\max}, \quad i \in \mathcal{A}(t), \quad (2.37d)$$

$$q_i = 0, \quad i \in \mathcal{A}^c(t). \quad (2.37e)$$

The optimization program can be viewed as a sequence of linear feasibility problems because constraint set is non-convex. We solve this program by an iterative approach. We start with a low easily achievable target rate  $r_0$  so that the constraint set (2.37b)-(2.37e) is feasible. We increase the target rate in small increments until the constraint set becomes infeasible. In each step, we get a set of feasible power assignments. The last set of feasible power assignments is the solution to the optimization program. The method converges, since the iterations generate a bounded sequence of increasing rates.

### 2.7.3 User discarding methodology

In our simulations, we eliminate 10% of users (over a large number of user realizations) in the following way. The procedure is identical for both the baseline and the system with relays. We simply assume the same peak power constraints for all base stations across the network. We fix the peak power threshold  $p_{\max}$  for each base station. Consider a single instantiation, where there are 57 users in the system. We increase the transmit power in all base stations in small incremental steps to improve the rate of the users in the system. As we go along, we discard the user associated with the base station whose power constraint goes active first. The base station is also turned off. This reduces the interference coupled with other users. Within the remaining set of

users, we can increase the transmit powers further. We then discard the next user causing the power constraint to go active and continue in this fashion until all remaining users in the system are guaranteed the desired rate of  $r_0$  bits/sec/Hz, without violating the peak power constraints. This procedure is repeated for a large number  $K$  of users instantiations. Let the number of users in outage for the  $k^{th}$  instantiation when the power threshold is  $p_{\max}$  and desired common rate  $r_0$  be  $O_k(p_{\max}, r_0)$ . We then find the total number of outage users for  $K$  instantiations when the peak power threshold is  $p_{\max}$  and the desired rate is  $r_0$  is calculated as

$$O(p_{\max}, r_0) = \sum_{k=1}^K O_k(p_{\max}, r_0). \quad (2.38)$$

The percentage of users in outage for the threshold  $p_{\max}$  and desired rate  $r_0$  is then

$$\frac{O(p_{\max}, r_0)}{57K} \times 100 \% \quad (2.39)$$

If  $O(p_{\max}, r_0) > 10\%$ , we increase the power threshold to  $p'_{\max} > p_{\max}$ . On the other hand, if  $O(p_{\max}, r_0) < 10\%$ , we decrease the power threshold to  $p''_{\max} < p_{\max}$ . Proceeding in this fashion, the base station peak power threshold  $p_{\max}$  is adjusted such that exactly 90% of the users are guaranteed the desired rate of  $r_0$  bits/sec/Hz and the rest of the 10% users are in outage.

The coordinates of the discarded users are stored and the same set of users are discarded when relays are present in the system too. We remark here that the order in which the users are discarded results in different power levels from the base stations, due to variations in the interference coupling among the users. Hence depending on the peak powers limitations at the base stations, the order in which the users are dismissed should be chosen carefully.

## 2.7.4 Network Operation and Simulation Aspects

### Baseline operation

We operate the baseline system as well as the system with relays such that, over a large number of user loading iterations, 90% of users obtain a common average rate of 1 bit/sec/Hz. We follow the approach described in Section 2.7.3 to discard users in the system. For the PC-CPA baseline, we solve a series of linear feasibility problem to obtain the base station powers for guaranteeing the desired common rate. One after another, we discard users that would cause the peak power constraint to go active. Hence we find the feasible set of powers  $p_1, \dots, p_N$  for the baseline such that 90% of the users get exactly 1 bit/sec/Hz.

### PC-CPA with relays: Average power savings

The peak power constraint of the base stations are fixed at  $p_{\max}$  such that the baseline can deliver 1 bit/sec/Hz at 10% outage. Since the relays are assumed to be inexpensive, we assume small peak power constraints for the relays. In our work, the peak powers of the relays are fixed at 1 W. Let us consider a single instantiation of 57 users in the system. We start off similar to the baseline after discarding the same set of users. The base stations target the users to deliver the common rate of 1 bit/sec/Hz. Only relays that have better SINR to the base stations than the user are eligible to help the user. The other relays are always inactive. At time  $t = 0$ , all relays are inactive. The aim in this experiment is to maintain a constant rate of 1 bit/sec/Hz throughout the course of the simulation. When relay  $i$  is eligible to transmit at time  $t$ , we include relay  $i$  into the set of active relays  $\mathcal{A}(t)$  and solve the LP (2.29). We stop when all the eligible relays are included in the set of active relays. The total power in the system when all eligible relays are active is noted down for this instantiation. We repeat this experiment for all the  $K$  instantiations.

### PC-CPA with relays: Peak power savings

The peak power constraints of the base stations are fixed at a value smaller than the baseline, say  $p'_{\max}$ . We assume inexpensive relays being deployed in the system. Thus, the peak power constraints of the relays are fixed at 1 W. Since the peak power value of the base stations is reduced from the baseline and the common rate is fixed at 1 bit/sec/Hz, the outage will be more than 10%. Let us consider a single instantiation of 57 users in the system. We start off similar to the baseline after discarding the same set of users. The base stations target the users to deliver the common rate of 1 bit/sec/Hz. Only relays that have better SINR to the base stations than the user are eligible to help the user. The other relays are always inactive. At time  $t = 0$ , all relays are inactive. Let relay  $i$  be eligible to transmit at time  $t$ . We include relay  $i$  into the set of active relays  $\mathcal{A}(t)$  and solve the LP (2.35). We stop when all the eligible relays are included in the set of active relays. We repeat this experiment for all the  $K$  instantiations and the outage is calculated. If the outage is less than 10%, the peak power of base stations is reduced to  $p''_{\max} < p'_{\max}$ , else the peak power values of the base stations is increased to  $p''_{\max} > p'_{\max}$  and the above procedure is repeated until the outage is close to 10%.

### 2.7.5 Simulation Results

#### Power savings

We observe that the average power savings averaged over  $K = 200$  instantiations are 3 dB. We observe that the peak power savings in the downlink when power control is employed is close to 2.6 dB.

#### Rate gains

We also performed experiments in which the rate improvements (problem (2.37)), and we observed 34% improvement in the throughput for 90% users in the system, with the baseline system being served at 1 bit/sec/Hz. The results are summarized in the Table 2.4.

Table 2.4: PC-CPA based relaying. Base station and relays employ power control

Peak power required to guarantee 1 bps/Hz at 10% outage Baseline (No relays)	Peak power required to guarantee 1 bps/Hz at 10% outage With relays	Savings in dB
10 W	5.5 W	2.6
Common rate for 90% users Baseline (No relays)	Common rate for 90% users With relays	Percentage rate increase
1 bps/Hz	1.34 bps/Hz	34%

## 2.8 Orthogonal Relaying

In the previous sections where we studied relaying with the CPA scheme, we saw that in an interference-limited setting, the improvement in throughput was limited by the interference due to multiple transmissions in the cell. In some cases the transmission from the base stations are redundant. For instance, for a user located at the edge of the cell, the received power from the base station could be weak and the base station's signal could be of little use. In that case, it might be better to turn the base station off since it could benefit the system overall in terms of reducing the interference levels. Moreover, the practical implementation of collaborative schemes can be complex with the existing technology. Hence, we investigate how much gains due to collaborative addition can be obtained if we just do simple multihopping, where the base station transmits to the relay in one slot and then the relay passes on the message to the destination in the next time slot. In this section, we exploit the half-duplex property of the relays in downlink cellular system to stagger the transmissions of the base station and relays over two time slots. A natural way to operate these relays is to have them receive in one time slot and transmit in another time slot. This gives us a natural orthogonality in the transmission scheme. Henceforth, we term this scheme as *orthogonal relaying*.

### 2.8.1 Network Operation and Simulation Aspects

The simulation set up is the same as that described in Section 2.3. Unlike CPA schemes, where the relay can start transmitting immediately after it decodes the message, relays can start transmitting only at specific times in the orthogonal relaying scheme. The system is assumed to be synchronous and time is divided into equal slots. The baseline and the system with relays are operated as follows.

#### Baseline

The baseline system operates similar to the P-CPA baseline as described in Section 2.6.1. All base stations transmit with peak powers and the users are required to get a fixed sized file with a specified deadline. Satisfied users leave the system as soon as they get the file. The associated base station sector is turned off. The users that do not get the file are in *outage*. In the learning phase, we learn the threshold peak transmit powers required by the base stations such that 10% of the users over a large population of users are in outage. The users in outage are discarded and the simulations are rerun.

#### System with relays

When relays are present in the system, time is divided into slots of equal durations, which is half of the baseline system. The operation of the system is periodic with odd and even time slots recurring at regular intervals. The base stations transmit in the odd time slots and the relays transmit in the even time slots. The peak power of the base stations are fixed as in the baseline and the peak power of the relays are fixed as 1 W.

- In the odd time slots, the base stations transmit at peak power. Relays are in receive mode in this time slot. The users and relays in each sector accumulate mutual information, depending on their channel qualities. If some of the users get the desired rate from the base station transmissions itself, those users are satisfied users and leave the system as soon as they get the desired rate. The corresponding base stations and relays are turned off. Let us denote the  $57 \times 1$

vector of rates obtained by users in the odd time slots by  $\mathbf{r}_o$ .

- In the even time slots, only the users that are yet to get the desired rate of 1 bps/Hz remain in the system. The base stations are turned off in this time slot. The relays that are required to help the users start transmitting simultaneously at the beginning of the even time slots. Simple power control is employed at the relay locations to reduce the interference caused to the other sectors. The power control is performed to achieve the desired residual common rate  $\mathbf{r}_e = \mathbf{1} - \mathbf{r}_o$ , where  $\mathbf{1}$  is a  $57 \times 1$  vector all 1's vector (representing 1 bps/Hz desired common rate). The users that require the relays to transmit more than their peak power constraint are discarded at the beginning of the even time slots. There may be cases where the user has a better channel to the base station than to the relay. Such users are not given the benefit of receiving the complete message from the base station. The base stations are switched off on the even time slots and are in outage if they do not get the message at the end of the even time slots.

## 2.8.2 User discarding method

### Baseline

Since the baseline scheme is the same as the P-CPA baseline, we follow the user discarding methodology as described in 2.6.2. In the initial learning phase, we find the power threshold such that 10% of the users that remain in the system are in outage. We then discard the users in outage and run the real network.

### System with relays

The user elimination procedure is same as that explained in the PC-CPA relaying scheme as described in Section 2.7.3. The users that violate the peak power constraint of the relays are discarded at the beginning of the even time slots. The discarded users do not get the desired common rate at the end of odd and even time slots, and hence are said to be in outage. The system is operated such that there is 10% outage in the system over a large number of user instantiations. The peak power threshold of the

relay nodes are adjusted such that there the outage percentage is exactly 10%.

For the real network simulations, we discard the 10% using the previously learnt power threshold, in the beginning of the odd time slot itself. We then run the network simulations as described in the previous Section 2.8.1.

### 2.8.3 Simulation Results

The average power savings in the base station locations is 3 dB, since the base stations transmit only for half the time. There is no peak power savings since the base stations transmit at peak power in the odd time slots. We obtain 35% rate gains due to orthogonal relaying when there is 10% outage in the system. It is interesting to note that simpler relaying methods, such as orthogonal relaying do nearly as well as the more complex forms of relaying, such as CPA schemes, in obtaining throughput gains and power savings. This observation is in agreement with the studies in simple linear settings [24, 32].

Table 2.5: SUMMARY OF GAINS DUE TO RELAYING: 19 cells, 57 sector network, one user per sector

Baseline (No Relay)	System with one relay per sector	Relay Benefits	
Baseline at 1 bps/Hz	Relay peak power = 1 W	Power Savings or Common Rate increase	
Peak power transmissions by base stations. Base peak power = 21 W	Peak power transmissions by base stations and relays (CPA)	1.46 dB (peak)	25%
Base station power control Base peak power = 10 W	Base station and relays power control (PC-CPA)	2.6 dB (peak) 3 dB (average)	34%
Peak power transmissions by base stations. Base peak power = 21 W	Relays power control to users (Orthogonal relaying)	3 dB (average)	35%

## 2.9 Conclusions

In this chapter, we presented a simulation study of the downlink of cellular system with relays. We evaluated the power savings and common rate increase for users when a common rate of 1 bps/Hz is required by 90% users in the system. We first described the collaborative power addition (CPA) scheme of relay collaboration. In the CPA based scheme, whenever the relay gets the complete message from the base station, it collaborates with the base station such that the mutual information at the user location corresponds to the sum of the received power at the user location, thus boosting up the average rate. We observe that when the system is interference limited the peak power savings are hard to come by. Consequently, the power control based collaborative power addition (PC-CPA) scheme along with a framework for power control is proposed. The power control framework can be posed as a linear program formulation when the objective is to minimize peak power or to minimize average energy in the system. This formulation can be used to evaluate the average and peak power savings in the system. The peak power savings and the rate gains improve when power control is employed. We then evaluated a simple multihopping scheme where the base stations and the relays transmit in orthogonal time slots. In the odd time slots, the base stations transmit at peak power and in the even time slots, the base stations are turned off and the relay employ simple power control to deliver the residual rate to the users. The summary of the results are given in the Table 2.5.

## Chapter 3

# Fair and Efficient Scheduling of Variable Rate Links via a Spectrum Server

### 3.1 Introduction

The previous chapter dealt with issues due to interference in cellular systems operating in licensed band. However, the emergence of unlicensed spectrum has spawned an impressive variety of important technologies and innovative uses, ranging from scientific and industrial to domestic applications and systems. Since these systems must adapt to a wide variety of unpredictable conditions, the emerging technologies called *cognitive radio* offer significant potential benefits in system capacity and service quality [33].

In their simplest embodiments (which are by no means simple to implement) cognitive radios can recognize the available systems and adjust their frequencies, waveforms and protocols to access those systems efficiently. Not surprisingly, it is upon these difficult “design” issues that most current research activities are focused. While these basic capabilities represent a difficult and significant step forward, they fail to fully illuminate the potential of cognitive behavior. They are perhaps analogous to a traveler with fluency in a variety of languages. Such fluency is great advantage, but how much greater an advantage is conferred when the traveler understands local conditions and customs, can choose the best language when several are possible, and can find local advisors and information when necessary. Following this analogy, cognitive radios must do more than communicate with the “local population” on an ad hoc basis to realize its full potential — it must develop a full awareness of a local environment that may span multiple spectrum bands and systems. This implies new discovery processes that are thorough and efficient, and even new classes of *information servers* that provide assistance in the process.

When there exist methods by which cognitive radios can independently discover local information, a variety of physical layer, system and network layer protocols can be applied to allow cooperation and coexistence [34]. However, such levels of cooperation and interoperability may not be possible when multiple services and systems must coexist. In a heterogeneous environment, some users may look to obtain high data rates without regard to energy efficiency; other users may wish to transmit at a fixed rate but with high efficiency. In certain applications, it will be important to enforce fairness constraints. In general, the system performance will have a multidimensional characterization. These dimensions represent conflicting performance measures.

In the realm of cognitive radio networks, two distinct sets of issues emerge. First, for a given transmitter and receiver technology and a specified set of performance measures, one must resolve the multidimensional boundaries of system performance. As we shall see, this is a difficult problem even if complete system state information is available to all network nodes. However, in any practical setting, complete information is not available at every node. In fact, cognitive radios must rely on some combination of individual measurement and shared information dispersed through the network. This information will be used for intelligent adaptation by the individual nodes as part of a large distributed system for spectrum allocation. This introduces a second set of issues relating to how the information distribution and link adaptation methods should be designed. A given set of distributed information gathering and exchange mechanisms may greatly influence the performance of the system.

In this chapter, we examine the boundaries of system performance under the assumption that efficient open access to spectrum can be resolved by impartial *spectrum servers* [6,35] that can obtain information about the interference environment through measurements contributed by different terminals, and then offer suggestions for efficient coordination to interested service subscribers. As observed in [6], likely neighborhood information could include various levels of time and frequency utilization, descriptions of nodes in a neighborhood, and potentially, spatial positions as well. In fact, the role of such a spectrum server for wireless network coordination is reminiscent of the role of the DHCP (Dynamic Host Configuration Protocol) server in the coordination problem

that arises among nodes in the Internet.

There are many different ways in which the spectrum server can coordinate a set of radios in a wireless network. While the work in [36] considers the role of the spectrum server in demand responsive pricing and competitive spectrum allocation, our work considers the role of the spectrum server in scheduling wireless links in an interference limited network. The links can have variable rates due to the variety of physical layer adopted by the links in the system.

### 3.1.1 Related Work

Scheduling transmissions in a wireless network has been studied in various contexts. Unlike the protocol model [37], where interference is modeled as a contention constraint [38], our model takes the physical layer into consideration. Although low-complexity scheduling algorithms [39,40] are still hot topics of research, the scheduling problem has been considered to jointly optimize another resource in the network. In [41], a joint scheduling and power control strategy is proposed to maximize network throughput and energy efficiency of the system. Their algorithm selects candidate subsets of concurrently active links, and applies the distributed power control algorithm [42] to find the minimal power vector. Another direction in this problem is addressed in [43,44], where the authors look at the cross-layer issues of routing, scheduling and power control. Joint scheduling and congestion control has been dealt with in [45,46]. The authors in [47] introduce the concept of transmission modes and develop a framework for integrated link scheduling and power control policies to maximize the average network throughput, when each link is subject to an average power constraint and each node is subject to a peak power constraint. The authors assume a model in which the data rate of a link is a linear function of the signal-to-interference ratio at the receiver. Most of the work in link scheduling in single hop networks or flow scheduling in multihop networks use mathematical programming approaches. For a recent survey of the research in this area, see [48].

### 3.2 System Model

It is recognized that the nodes of a cognitive radio network can interact in a wide variety of (arbitrary) ways. To distill these interactions, we observe that each radio follows a transmission policy that results in signals that vary over time, frequency, and space. This variation may be the result of adaptation to measurements of channels or interference. The performance of a particular signaling strategy depends on each receiver's ability to resolve signals in the presence of interfering transmissions.

For a system of interfering wireless transmissions, a mathematical model starts with a basis for the signal space. Conceptually, this is achieved by K-L decomposition. Given a basis, each user transmission employs a combination of basis functions and transmits in some or all of the signals dimensions. An important lesson from communication theory is that the actual system performance does not depend on this basis, but rather the choice of basis determines how well we can understand the communication system.

This work assumes a relatively simple signaling structure. We assume that the signals can be decomposed into time-slotted narrowband transmissions. Within a time slot, the transmitters use variable-rate coding to combat interference and/or to coexist with other transmitters. The interference that a receiver faces depends on the subset of nodes transmitting in that time slot. We assume that the data rate obtained by each user can be decomposed into the data rate obtained on each signal dimension. Moreover, for the average data rate to be a meaningful, the collection of user policies must result in an ergodic signaling process. With respect to a signal dimension, the number of bits that can be transmitted on a link depends on the ability of the receiver to separate the desired signal from interfering signals. Such an ability is limited by the technology of the receiver.

Let us consider a wireless network with  $N$  nodes forming  $L$  point-to-point links sharing a common spectrum. We model the network as a directed graph  $\mathcal{G}(\mathcal{V}, \mathcal{E})$ , where the nodes in the network are represented by the set of vertices  $\mathcal{V}$  of the graph and the links are represented by a set of directed edges  $\mathcal{E}$ . A directed edge from a node  $m$  to node  $n$  implies that  $n$  wishes to communicate data to node  $m$ . We study the scenario

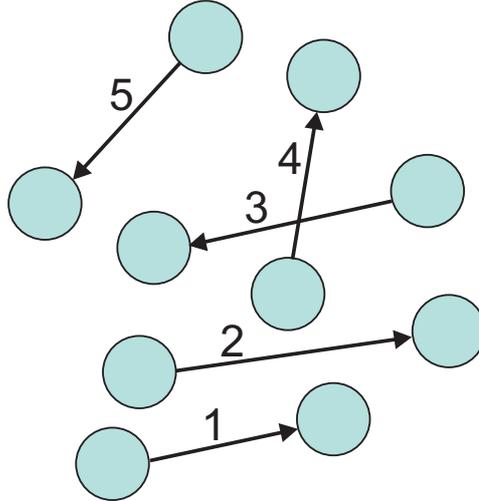


Figure 3.1: Graph of network showing the nodes and directed links

where the spectrum server coordinates the activity of the set of  $L$  links to share the spectrum efficiently.

Before we explain the system model, we comment on the notation of this chapter. We use boldface lowercase characters for vectors and boldface uppercase for matrices. If  $\mathbf{a}$  is a vector,  $\mathbf{a}^T$  denotes its transpose and  $\mathbf{a}^T \mathbf{b} = \sum_i a_i b_i$  represents the inner product of the vectors  $\mathbf{a}$  and  $\mathbf{b}$ . The vector of all zeros and all ones are represented by  $\mathbf{0}$  and  $\mathbf{1}$  respectively.

Define the set of *transmission modes*  $\mathcal{T} = \{0, 1, \dots, M - 1\}$ , where  $M$  denotes the number of possible transmission modes. Then the *mode activity vector*  $t_i$  of mode  $i$  is a binary vector, indicating the on-off activity of the links. If  $t_i = (t_{1i}, t_{2i}, \dots, t_{Li})$  is a mode activity vector, then

$$t_{li} = \begin{cases} 1, & \text{link } l \text{ is active under transmission mode } i, \\ 0, & \text{otherwise.} \end{cases} \quad (3.1)$$

Note that  $M$  is number of available transmission modes including the mode in which all links are off. For a given system of links, some of the modes may be infeasible due to physical constraints on the system or due to the constraints placed by the network. Figure 3.1 shows a representative network and Figure 3.2 shows particular transmission

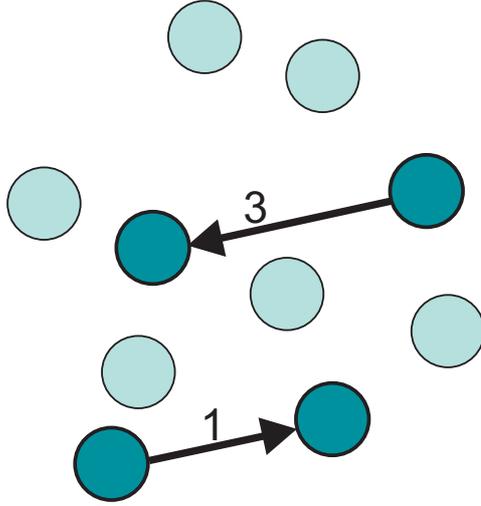


Figure 3.2: Graph of network showing transmission mode corresponding to  $[1 \ 0 \ 1 \ 0 \ 0]$

mode for the set of links.

We consider transmission scenarios in which the rate in any link is also determined by transmissions by other links in that particular mode. We denote the rate obtained in the link  $l$  under mode  $i \in \mathcal{T}$  as  $c_{li}$ . Let  $x_i$  be the fraction of time that transmission mode  $i$  is active. We refer to the vector  $\mathbf{x} = [x_0 \ x_1 \ \dots \ x_M]^T$  as a schedule without precise specification of sequence of active modes. The average data rate in link  $l$  is the time average of the data rates of all the transmission modes that include link  $l$  and is given by

$$r_l = \sum_i c_{li} x_i, \quad (3.2)$$

or in vector form,

$$\mathbf{r} = \mathbf{C}\mathbf{x}, \quad (3.3)$$

where  $\mathbf{C}$  is a  $L \times M$  matrix with non-negative entries  $[\mathbf{C}]_{li} = c_{li}$ ,  $\mathbf{r}$  is a real vector of length  $L$  and  $\mathbf{x}$  is a real vector of length  $M$ . Embedded in the matrix  $\mathbf{C}$  are the rates obtained in each link  $l$  as a part of transmission mode with simultaneous transmissions on multiple links.

We note that this model allows for consideration of a large class of physical layer interactions. We observe that all necessary aspects of transmitter and receiver technology

are embedded in the rate matrix  $\mathbf{C}$ . In what follows, we provide two specific examples of how the matrix  $\mathbf{C}$  can be constructed to encompass various multiuser communication strategies.

### Gaussian Interference Channel

Let the transmit power on a link  $l \in \mathcal{E}$  be  $P_l$ . If  $G_{lk}$  is the link gain from the transmitter of link  $k$  to the receiver of link  $l$  and  $\sigma_l^2$  is the noise power at the receiver of link  $l$ , the Signal-to-Interference plus Noise Ratio (SINR)  $\gamma_{li}$  at the receiver of link  $l$  in transmission mode  $i$  is given by

$$\gamma_{li} = \frac{t_{li}G_{ll}P_l}{\sum_{k \in \mathcal{E}, k \neq l} t_{ki}G_{lk}P_k + \sigma_l^2}. \quad (3.4)$$

The link gain between a transmitter and receiver takes into account the path loss and attenuation due to shadow fading. We assume that the link gains between each transmitter and receiver are known to the spectrum server. The data rate in each link depends on the SINR in that link. We assume that the transmitter can vary its data rate, possibly through a combination of adaptive modulation and coding. In particular, for a given mode, the transmitter and receiver on a link employ the highest rate that permits reliable communication given the link SINR in that mode. If we assume that the transmission of other links are treated as Gaussian noise and that a transmission on link  $l$  is reliable in a given mode  $i$  with a data rate (logarithms are to the base 2)

$$c_{li} = \log(1 + \gamma_{li}). \quad (3.5)$$

For sake of simplicity, we do not consider any minimum SINR threshold required at each receiver, i.e., associated with each transmission mode  $i$ , a non-zero  $\gamma_{li}$  defines some rate on the link  $l$ . Let  $x_i$  be the fraction of time that transmission mode  $i$  is active and  $r_l$  be the average data rate of link  $l$ . The average data rate in link  $l$  is the time average of the data rates of all the transmission modes that include link  $l$ . Thus,

$$r_l = \sum_i c_{li}x_i, \quad (3.6)$$

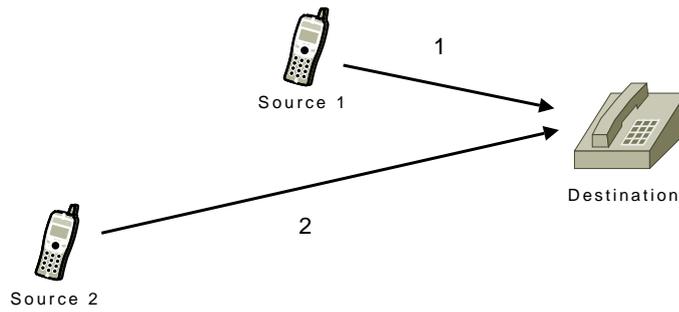


Figure 3.3: Multiaccess Channel

or in vector form,

$$\mathbf{r} = \mathbf{C}\mathbf{x}, \quad (3.7)$$

where  $\mathbf{C} = [\mathbf{c}_1 \ \mathbf{c}_2 \ \dots \ \mathbf{c}_M]$  is an  $L \times M$  matrix with non-negative entries, such that its  $j$ th column  $\mathbf{c}_j = [c_{1j}, c_{2j}, \dots, c_{Lj}]^T$  contains the rate obtained by each link in mode  $j$ . We denote  $\mathbf{P} = \text{diag}(P_1, P_2, \dots, P_L)$  to be the  $L \times L$  diagonal matrix of transmit powers of the individual links. Let  $\mathbf{T} = [\mathbf{t}_1 \ \mathbf{t}_2 \ \dots \ \mathbf{t}_M]$  denote the  $L \times M$  binary matrix, which contains the transmission mode activity vector  $\mathbf{t}_j$  as column  $j$ .

### Gaussian Multiple Access Channel

Figure 3.3 shows a multiple access scenario wherein two terminals send independent information to a receiver. Assume that the receiver knows the channel gains between transmitters to itself, and that the transmitters encode their data using a capacity-achieving channel code. The receiver could then perform successive decoding to achieve a rate pair  $(r_1, r_2)$  for the links 1 and 2. Figure 3.4 shows the set of achievable rate pairs that can be achieved by any successive decoding scheme [25, Chapter 15]. In particular, for the above multiple access scenario, let  $C_1 = (a, 0)$  and  $C_2 = (0, b)$  are the rate pairs that can be achieved if the links transmit in isolation. If the link 1 is decoded first and its interference on link 2 is removed, the rate pair  $C_4 = (c, b)$  is achievable. On the other hand, if link 2 is decoded first to cancel out its interference on the link 1, the rate pair  $C_3 = (a, d)$  is achievable.

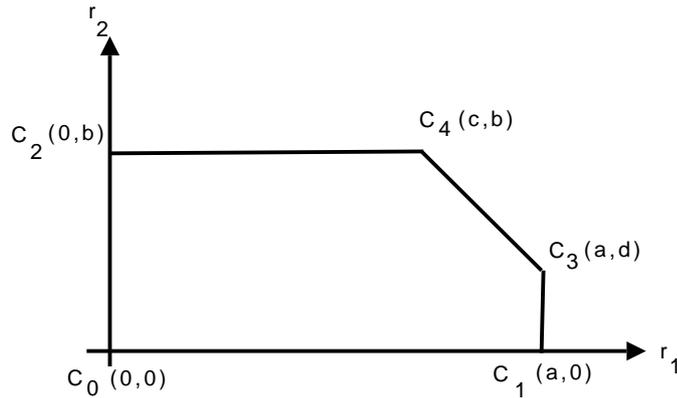


Figure 3.4: Rate region of the multiaccess channel

Thus, the rate matrix  $\mathbf{C}_{MA}$  is given by,

$$\mathbf{C}_{MA} = \begin{bmatrix} 0 & a & 0 & c & a \\ 0 & 0 & b & b & d \end{bmatrix}. \quad (3.8)$$

Any rate pair that can be achieved over the multiaccess channel is a convex combination (achieved by time-sharing) of the rate pairs corresponding to the points  $C_0, C_1, C_2, C_3, C_4$  of the Figure 3.3.

In summary, we note that this model has a number of desirable characteristics. First, as illustrated by the above examples, we observe that all aspects of transmitter and receiver technology are embedded in the rate matrix  $\mathbf{C}$ . If, for instance, the links employed CDMA spreading, (3.4) for the SINR on link  $l$  in mode  $j$  would be appropriately modified, as in [3], to reflect the transmitter spreading sequences and receiver filter vectors used in that mode. Similarly, if we were to assume a particular practical coding scheme, we would modify Equation (3.5) for the expected number of bits that we would expect to decode at a specified SINR. Thus the general model allows for consideration of a large class of physical layer interactions. We employ the specific choices in Equations (3.4) and (3.5) to demonstrate trade-offs in between average rates and various fairness constraints.

The set of transmission modes  $\mathcal{T}$  may turn out to be uncountable for some transmission schemes. For example, if the nodes of the network can use a continuum of powers at their transmitters, each power level gives rise to a transmission mode. However, we could ‘discretize’ the set of transmission modes by allowing appropriately chosen finite power levels.

In practice, a scheduler will specify a sequence of transmission modes. Typically, this would be done by constructing a frame with  $N$  time slots and allocating  $N_j$  time slots to each mode  $j$ . The fraction of time that mode  $j$  is active will be  $x_j = N_j/N$ . For sufficiently large  $N$ , the ratio  $N_j/N$  can be made arbitrarily close to any  $x_j \in [0, 1]$ . In this case, the average rate  $\mathbf{r}$  in (3.3) will represent the average link data rates over one frame. For our analytical model, we optimize these average rates per frame by specification of the time fractions in  $\mathbf{x}$ , without explicitly specifying the precise slots assigned to each mode.

Conversely, consider the average link rates obtained by an arbitrary dynamic spectrum access system. Each link employs a dynamic policy, based on measurements and perhaps some side information, to determine when to be active. At any given time, some subset of links will be active and the rates obtained on each link will be determined by the interference generated by those active links. In short, any dynamic spectrum access system yields a series of transmission modes. The rate obtained by each link  $l$  in each mode  $j$  will be given by  $c_{lj}$ . To speak of average rates for the links, the collection of link access policies must yield an ergodic transmission mode process such that we can define  $x_j$  as the fraction of time the system is in mode  $j$ . In this case, the average link data rates will be given by (3.3). In short, any set of average rates obtained by a dynamic spectrum access system can be obtained by a centralized scheduler that specifies the identical time fraction  $x_j$  for each mode  $j$ . The centralized scheduler allows us to separate what average link rates can be obtained from the issue of whether a dynamic system can achieve those rates.

### 3.3 Maximum Sum Rate Scheduling

The centralized approach allows us to optimize global objective functions, i.e., objective functions that allow us to optimize the overall system performance, which is difficult to optimize in a distributed setting. In this section, we are interested in the schedule that maximizes the sum of the average data rates over all links  $l = 1, 2, \dots, L$ , subject to constraints on the minimum rates for each link. These kinds of objective functions are useful when the overall utility of the network needs to be maximized, while provided minimum guarantees to the end users. Let us assume that each link has a minimum average data rate requirement  $r_{\min,l}$ . The optimization problem for finding the maximum sum rate schedule can be posed as the following linear program (LP):

$$\max_{\mathbf{r}, \mathbf{x}} \quad \mathbf{1}^T \mathbf{r} \quad (3.9a)$$

$$\text{subject to} \quad \mathbf{r} = \mathbf{C}\mathbf{x}, \quad (3.9b)$$

$$\mathbf{r} \geq \mathbf{r}_{\min}, \quad (3.9c)$$

$$\mathbf{1}^T \mathbf{x} = 1, \quad (3.9d)$$

$$\mathbf{x} \geq \mathbf{0}. \quad (3.9e)$$

The objective function  $\mathbf{1}^T \mathbf{r} = \sum_i r_i$  is the sum of average rates of the individual links. The inequality in (3.9c) represents the minimum rate constraint and (3.9d) is the normalization for the schedule. The variables in the LP (3.9) are  $\mathbf{r}$  and  $\mathbf{x}$ . Rewriting the LP in terms of the variable  $\mathbf{x}$  only, we get

$$c_{\text{opt}}(\mathbf{r}_{\min}) = \max_{\mathbf{x}} \quad \mathbf{1}^T \mathbf{C}\mathbf{x} \quad (3.10a)$$

$$\text{subject to} \quad \mathbf{C}\mathbf{x} \geq \mathbf{r}_{\min}, \quad (3.10b)$$

$$\mathbf{1}^T \mathbf{x} \leq 1, \quad (3.10c)$$

$$\mathbf{x} \geq \mathbf{0}. \quad (3.10d)$$

Since  $\mathbf{C}$  is a matrix with non-negative entries, the constraint  $\mathbf{1}^T \mathbf{x} = 1$  can be replaced by the constraint  $\mathbf{1}^T \mathbf{x} \leq 1$  since the optimum  $\mathbf{x}$ , say  $\mathbf{x}_{\text{opt}}$ , will satisfy  $\mathbf{1}^T \mathbf{x}_{\text{opt}} = 1$ .

Otherwise, we could scale  $\mathbf{x}_{\text{opt}}$  up so that the objective function is increased because the objective function is non-negative. We denote the optimal value  $\mathbf{1}^T \mathbf{C} \mathbf{x}_{\text{opt}}$  as  $c_{\text{opt}}(\mathbf{r}_{\text{min}})$ .

### 3.3.1 No minimum rate constraint

We now consider the special case when  $\mathbf{r}_{\text{min}} = \mathbf{0}$ , i.e., when there is no minimum rate requirement for any of the links.

**Theorem 1** *When  $\mathbf{r}_{\text{min}} = \mathbf{0}$ , the solution to the LP (3.10) is  $\mathbf{x}_{\text{opt}} = [0 \ 0 \ \dots \ 1 \ \dots \ 0 \ 0]^T$ , where the position of 1 corresponds to the transmission mode with the maximum sum rate. The optimal objective value is the maximum column sum of the rate matrix  $\mathbf{C}$ . Thus, the optimal strategy is to always operate the transmission mode with the maximum sum rate.*

• **Proof:** Since  $\mathbf{C}$  is non-negative and  $\mathbf{r}_{\text{min}} = \mathbf{0}$  by hypothesis, any  $\mathbf{x}$  satisfying (3.10c) and (3.10d) is feasible, as (3.10b) is trivially satisfied. Since  $\mathbf{1}^T \mathbf{C}$  represents the row-vector of column sums of  $\mathbf{C}$ , the objective function  $\mathbf{1}^T \mathbf{C} \mathbf{x}$  is a convex combination of column sums of  $\mathbf{C}$ . Therefore,

$$\mathbf{1}^T \mathbf{C} \mathbf{x} = \sum_{l=1}^L \sum_{i=1}^M c_{li} x_i \quad (3.11)$$

$$= \sum_{i=1}^M x_i \sum_{l=1}^L c_{li} \quad (3.12)$$

$$\leq \sum_{i=1}^M x_i \max_i \sum_{l=1}^L c_{li} \quad (3.13)$$

$$= \max_i \sum_{l=1}^L c_{li} \quad (3.14)$$

where the equality in (3.14) is true since  $\sum_i x_i = 1$ . Equality holds in (3.13) when  $\mathbf{x} = \mathbf{x}_{\text{opt}} = [0 \ 0 \ \dots \ 1 \ \dots \ 0 \ 0]^T$  where the position of 1 in  $\mathbf{x}_{\text{opt}}$  is  $\hat{i} = \arg \max_i \sum_{l=1}^L c_{li}$ .

■

We refer to transmission mode with the maximum sum rate as the *dominant transmission mode*. Depending on the geometry of the links, the dominant transmission

mode can be a single active link or a collection of geographically separated links. However, one implication of the above theorem is that the links that are not part of the dominant transmission mode get zero rate. So, the system is unfair in terms of providing non-zero data rates to all the links.

### 3.3.2 Non-zero minimum rate constraint

In order to offset the unfairness in the system, we introduce a non-zero minimum rate requirement in the individual links of the network. In such cases, the optimal schedule balances the use of the dominant mode against modes that provide non-zero rates to the otherwise disadvantaged links. This, however, comes at the cost of reduction in the sum rate of the network. In the case when  $\mathbf{r}_{\min}$  is non-zero, there is an additional constraint in (3.10b) which has to be met. The optimal objective value cannot exceed the unconstrained optimum  $c_{\text{opt}}(\mathbf{0})$ .

We can characterize the loss in sum rate due to the minimum rate constraint. We begin by writing the dual problem for the LP. The Lagrangian for the LP (3.10) is

$$L(\mathbf{x}, \mathbf{u}, v) = \mathbf{1}^T \mathbf{C} \mathbf{x} + \mathbf{u}^T (\mathbf{C} \mathbf{x} - \mathbf{r}_{\min}) + v(1 - \mathbf{1}^T \mathbf{x}), \quad (3.15)$$

where  $\mathbf{u} \in \mathcal{R}_+^L$  and  $v \in \mathcal{R}_+$  are the Lagrange dual variables. The Lagrange dual function is

$$g(\mathbf{u}, v) = \sup_{\mathbf{x} \geq \mathbf{0}} L(\mathbf{x}, \mathbf{u}, v) \quad (3.16)$$

$$\begin{aligned} &= -\mathbf{u}^T \mathbf{r}_{\min} + v + \sup_{\mathbf{x} \geq \mathbf{0}} (\mathbf{1}^T \mathbf{C} + \mathbf{u}^T \mathbf{C} - v \mathbf{1}^T) \mathbf{x} \\ &= \begin{cases} -\mathbf{u}^T \mathbf{r}_{\min} + v, & \mathbf{1}^T \mathbf{C} + \mathbf{u}^T \mathbf{C} - v \mathbf{1}^T \leq \mathbf{0} \\ \infty, & \text{otherwise.} \end{cases} \end{aligned} \quad (3.17)$$

Thus the dual of the LP (3.10) is

$$\min_{\mathbf{u}, v} \quad -\mathbf{r}_{\min}^T \mathbf{u} + v \quad (3.18a)$$

$$\text{subject to} \quad \mathbf{C}^T(\mathbf{1} + \mathbf{u}) \leq v\mathbf{1}, \quad (3.18b)$$

$$\mathbf{u} \geq \mathbf{0}, v \geq 0. \quad (3.18c)$$

Since (3.18) is also an LP, by strong duality [49, Chapter 5], the optimal value of the dual problem in (3.18) is equal to  $c_{\text{opt}}(\mathbf{r}_{\min})$ . Let  $(\mathbf{u}^*, v^*)$  be the solution of (3.18). Since by Theorem 1,  $c_{\text{opt}}(\mathbf{0})$  is the maximum column sum of  $\mathbf{C}$ , i.e.,  $c_{\text{opt}}(\mathbf{0}) = \max \mathbf{C}^T \mathbf{1}$ , we have according to (3.18b),  $v^* \geq c_{\text{opt}}(\mathbf{0}) + \max \mathbf{C}^T \mathbf{u}^*$ . Therefore, the optimal value of (3.18)

$$\begin{aligned} c_{\text{opt}}(\mathbf{r}_{\min}) &= -\mathbf{r}_{\min}^T \mathbf{u}^* + v^* \\ &\geq -\mathbf{r}_{\min}^T \mathbf{u}^* + \max \mathbf{C}^T \mathbf{u}^* + c_{\text{opt}}(\mathbf{0}). \end{aligned}$$

Since  $c_{\text{opt}}(\mathbf{0}) - c_{\text{opt}}(\mathbf{r}_{\min}) \leq \mathbf{r}_{\min}^T \mathbf{u}^* - \max \mathbf{C}^T \mathbf{u}^*$ , the loss in sum rate is at most  $\mathbf{r}_{\min}^T \mathbf{u}^* - \max \mathbf{C}^T \mathbf{u}^*$ . For the simple case when  $\mathbf{r}_{\min} = r_{\min} \mathbf{1}$ , we have

$$\frac{c_{\text{opt}}(\mathbf{0}) - c_{\text{opt}}(r_{\min})}{r_{\min}} \leq \mathbf{1}^T \mathbf{u}^* - \max \frac{\mathbf{C}^T \mathbf{u}^*}{r_{\min}}. \quad (3.19)$$

Thus, the rate of change in the optimal value of the sum rate in the system can be upper bounded by the right hand side expression.

There exists a trade-off between the sum rate and individual rates of the links, i.e., when we increase the minimum rate requirement in the links, the sum rate obtained decreases. This is intuitively satisfying since the dominant mode, which offers the highest sum rate, is always turned on whenever there is no minimum rate requirement on the links. When the minimum rate requirement is increased from zero, other transmission modes are forced to be scheduled for transmission in order to satisfy the minimum rate requirement of the links. Since the modes other than the dominant mode always offer a lesser sum rate than the dominant mode, the sum rate decreases monotonically with

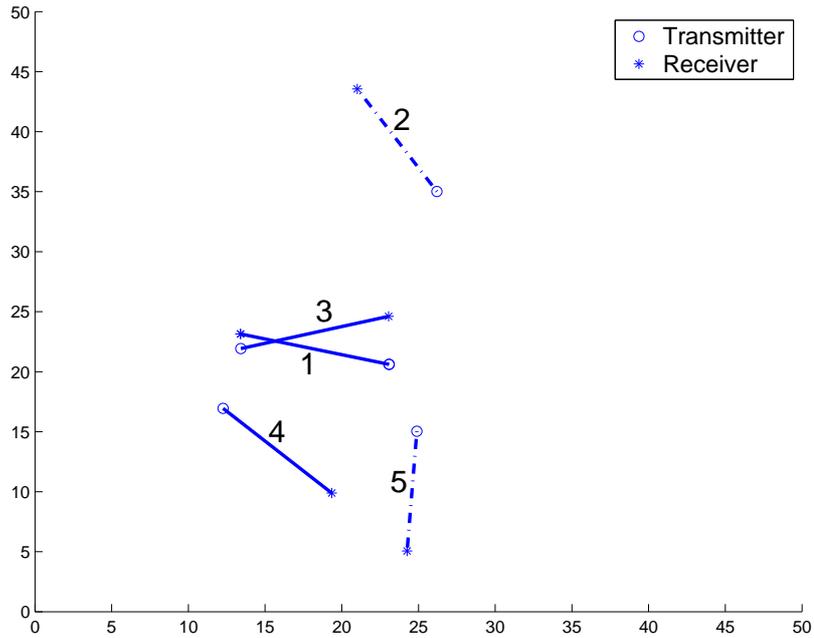


Figure 3.5: A set of source-destination pairs

increase in required minimum rate. The minimum rate requirement for all links in the network can be increased by trading off sum rate until it is infeasible to support the rate requirement in all links.

### 3.3.3 Simulation Results

We now consider an example scenario in Figure 3.5. The simulation set-up is a  $50 \times 50$  grid. We use this scenario to as a running example to illustrate the performance of various scheduling schemes in this chapter. We assume the Gaussian interference channel model described in Section 3.2. The links are of fixed lengths and placed at random locations in the grid. The interference gain  $G_{lj}$  between the transmitter of link  $j$  and the receiver of a link  $l$  is given by  $G_{lj} = d_{lj}^{-4}$ , where  $d_{lj}$  is the separation distance between the transmitter and receiver. The transmit powers are fixed for all transmissions and the link geometries are characterized through the signal-to-noise ratio (SNR) at the receiver for that link (in the absence of interference).

In the special case when the noise at the receiver is high, the denominator in the

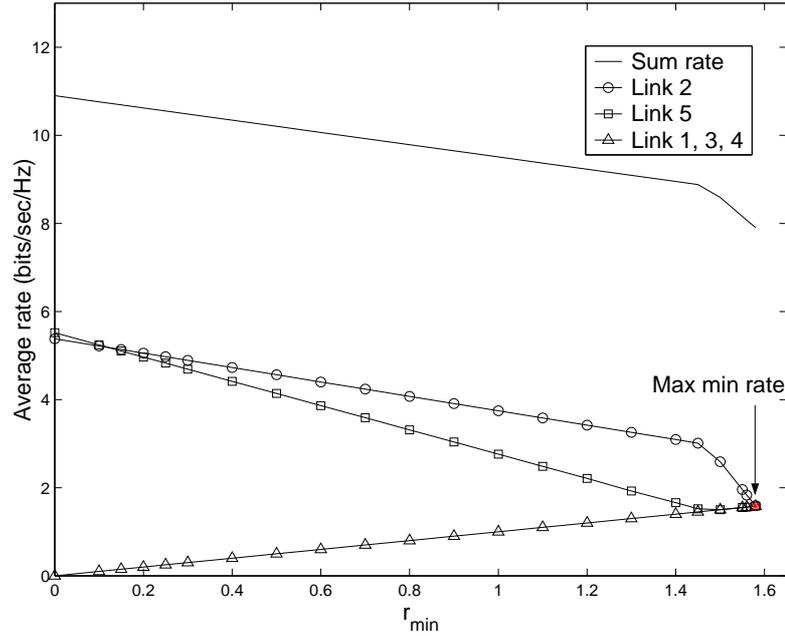


Figure 3.6: Variation of sum rates and individual rate as a function of  $r_{\min}$

SIR expression (3.4) is dominated by the receiver noise. This approximates the case when there is no interference from the neighboring links. Hence the best policy would be to turn on all the links in order to maximize the sum rate in all the links. However, in the case of high SNR links, the best policy is to operate the singleton modes.

As the SNR in each link increases, the interference from neighbors also increases. Then the best transmission mode is that which has the highest sum rate among all the other transmission modes. The set of links chosen follows spatial reuse patterns that are reminiscent of those used in cellular networks. Figure 3.5 shows a set of links and the dominant transmission mode at SNR = 20 dB. The links in the dominant mode are shown in solid lines.

In the case of maximum sum rate scheduling with non-zero minimum rate constraint, we see that more than one transmission mode is operated since there is a minimum rate requirement for each link. The dominant mode is selected for most of the time and the mode which includes the poorer quality links are turned on for a fraction of time just enough to satisfy their minimum rate requirement. For example with 20 dB link SNR in

the topology of Figure 3.5, the mode consisting of links  $\{2, 5\}$  is always operated when  $\mathbf{r}_{\min}$  is zero. But as the common minimum rate  $\mathbf{r}_{\min}$  increases from zero, an additional set of modes ( $\{2, 4\}, \{1\}, \{3\}$ ) is operated to satisfy the minimum rate requirement for each link. The schedules of the individual transmission modes varies with the minimum rate so that the minimum rate constraint in each of links is maintained. Notice that only five distinct modes are active. When  $\mathbf{r}_{\min}$  is increased in steps, we observe that the same set of modes ( $\{2, 5\}, \{2, 4\}, \{1\}, \{3\}$ ) is operated. Figure 3.6 shows the trade-off between the sum rate and the minimum rate requirement of the individual links. After a certain  $\mathbf{r}_{\min}$  value, say  $\mathbf{r}'$ , a different set of modes ( $\{2, 5\}, \{2, 4\}, \{1\}, \{3\}, \{4\}$ ) has to be operated in order to obtain a feasible schedule. Until then the sum rate falls linearly with increase in  $\mathbf{r}_{\min}$ . The break point in the sum rate curve occurs at  $\mathbf{r}' = 1.48$ .

### 3.3.4 Maximum sum rate schedule with high SNR links

We examine the special case of high SNR links when each link transmits with a large power  $P$  in the Gaussian interference model. Define a set of *singleton* modes

$$\hat{\mathcal{T}} = \{i_l : t_{li_l} = 1, t_{ki_l} = 0 \text{ for all } k \neq l\}.$$

In mode  $i_l$ , link  $l$  transmits in isolation and thus we call  $\hat{\mathcal{T}} = \{i_1, i_2, \dots, i_L\}$  the set of singleton modes. When the transmit power  $P$  is high, all links have high SNR and a link  $l$  achieves a high rate when transmitting in the singleton mode  $i_l$ . However, in a shared (non-isolation) mode  $j \notin \hat{\mathcal{T}}$ , links will have interference-limited SINRs and relatively low data rates. These observations lead to the following theorem.

**Theorem 2** *If the interference gains  $G_{lk}$  are all non-zero, then for sufficiently large transmit power  $P$ , the solution to (3.10) is time sharing among the transmission modes in  $\hat{\mathcal{T}}$ .*

• **Proof:** If  $P$  is the transmit power in all links  $l \in \mathcal{E}$ , from (3.4) the SIR  $\gamma_{lj}$  of link  $l$  in transmission mode  $j$  is given by

$$\gamma_{lj} = \frac{t_{lj}G_{ll}P}{\sum_{j \in \mathcal{E}, k \neq l} t_{kj}G_{lk}P + \sigma_l^2}. \quad (3.20)$$

For all modes  $j \notin \hat{\mathcal{T}}$ , the nonzero interference gains  $G_{lk}$  and the monotonicity of the fraction  $P/(cP + \sigma^2)$  imply that

$$\gamma_{lj} < \bar{\gamma}_{lj} = \frac{G_{ll}}{\sum_{j \in \mathcal{E}, k \neq l} t_{kj}G_{lk}}. \quad (3.21)$$

We can thus upper bound the SIR  $\gamma_{lj}$  of any link  $l$  in any transmission mode  $j \notin \hat{\mathcal{T}}$  as

$$\gamma_{lj} < \bar{\gamma} = \max_{j \notin \hat{\mathcal{T}}} \max_l \bar{\gamma}_{lj}. \quad (3.22)$$

It follows from (3.5) that

$$c_{lj} \leq \bar{c} = \log(1 + \bar{\gamma}), \quad j \notin \hat{\mathcal{T}}. \quad (3.23)$$

Note that  $\bar{c}$  serves as an upper bound for the rate that can be obtained by any link  $l$  in a shared mode  $j \notin \hat{\mathcal{T}}$ . However, in a mode  $i_l \in \hat{\mathcal{T}}$  in which only link  $l$  is active,

$$\gamma_{i_l} = \frac{G_{ll}P}{\sigma_l^2} = \gamma_l(P), \quad (3.24)$$

a monotone increasing function of  $P$ . Let us define

$$c_l(P) = \log(1 + \gamma_l(P)). \quad (3.25)$$

as the data rate obtained when link  $l$  transmits with power  $P$  in the singleton mode  $i_l$ . Since  $c_l(P)$  is a monotone increasing function of  $P$ , there exists a transmit power  $P^*$ , such that  $P > P^*$  implies  $c_l(P) > L\bar{c}$  for all links  $l$ .

Now, let us suppose that  $P > P^*$ , but  $\mathbf{x}$  is an optimal schedule for problem (3.10) with  $x_j > 0$  for a shared mode  $j \notin \hat{\mathcal{T}}$ . Consider a new schedule  $\mathbf{x}'$  given by

$$x'_i = \begin{cases} 0 & i = j \\ x_i + x_j/L & i \in \hat{\mathcal{T}} \\ x_i & \text{otherwise} \end{cases} \quad (3.26)$$

The schedule  $\mathbf{x}'$  reallocates the time  $x_j$  in mode  $j$  equally to the isolation modes  $i_l$  in  $\hat{\mathcal{T}}$ . In particular, an isolation mode  $i_l \in \hat{\mathcal{T}}$  will now be active for time

$$x'_{i_l} = x_{i_l} + \frac{x_j}{L}. \quad (3.27)$$

We now show that every link  $l$  receives a positive rate increase by switching to schedule  $\mathbf{x}'$ . Under schedule  $\mathbf{x}$ , a link  $l$  obtains rate

$$r_l = \sum_i c_{li}x_i = c_{lj}x_j + c_{li}x_{i_l} + \sum_{i \notin \{j, i_l\}} c_{li}x_i. \quad (3.28)$$

Under schedule  $\mathbf{x}'$ , link  $l$  obtains rate

$$r'_l = \sum_i c_{li}x'_i = c_{li}x'_{i_l} + \sum_{i \notin \{j, i_l\}} c_{li}x_i. \quad (3.29)$$

For link  $l$ , the difference in rates is

$$r'_l - r_l = c_{li}(x'_{i_l} - x_{i_l}) - c_{lj}x_j \quad (3.30)$$

$$= \left( \frac{c_{li}}{L} - c_{lj} \right) x_j. \quad (3.31)$$

However,  $P > P^*$  implies that in the isolation mode  $i_l$ , link  $l$  obtains rate

$$c_{li} = c_l(P) > L\bar{c}. \quad (3.32)$$

It follows that  $r'_l - r_l > 0$  for all links  $l$ . This contradicts the optimality of schedule  $\mathbf{x}$  in that every link achieves a strictly higher rate under schedule  $\mathbf{x}'$ .

### 3.4 Fair scheduling

In Section 3.3, we observed that maximizing sum rate without minimum rate constraints on the links leads to unfairness among the links. Usually, when global objective functions like sum rate in the network is involved, issues of fairness naturally arise. We investigate fair scheduling strategies in this section. We start by looking at the conventional max-min fair scheme.

• **Definition:** A vector of rates  $\mathbf{r}$  is said to be *max-min fair* if it is feasible and for each  $l \in \mathcal{E}$ ,  $r_l$  cannot be increased while maintaining feasibility without decreasing  $r_{l'}$  for some link  $l'$  for which  $r_{l'} \leq r_l$ . Formally, for any other feasible allocation  $\tilde{\mathbf{r}}$ , with  $\tilde{r}_l > r_l$ , there must exist some  $l'$  such that  $\tilde{r}_{l'} < r_{l'} \leq r_l$ .

Max-min fairness is well studied in the context of data networks [21], in the context of flow control of elastic traffic. The data network models differ from the one we consider since in the former, there are multiple flows through many links of finite capacity. There may be several bottleneck links and a feasible rate allocation is max-min fair if and only if all flows pass through at least one bottleneck link [21]. However, in our model, each link has a minimum rate requirement to satisfy and it is not clear as to what is the bottleneck in this case. Even if we could answer this question, the next important question is how many bottlenecks are there in the system? Given that the max-min fair set of rates exist, how do we compute them? We try to answer these questions in this section.

In order to obtain the max-min fair schedule in our setting, we begin by formulating the LP to maximize the minimum common rate in all the links. The LP which

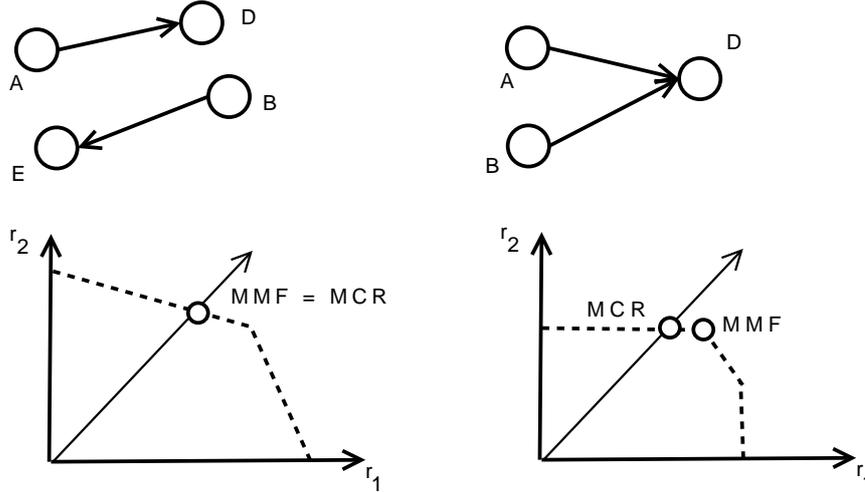


Figure 3.7: Illustration of the solidarity property. The maximizing the common rate leads to max-min fairness in the Gaussian interference model. Whereas in the multiple access model, maximum common rate and max-min fair rates are unequal.

maximizes the minimum common rate among the links is

$$r^* = \max_{\mathbf{r}, \mathbf{x}} r_{\min} \quad (3.33a)$$

$$\text{subject to } \mathbf{r} = \mathbf{C}\mathbf{x}, \quad (3.33b)$$

$$\mathbf{r} \geq r_{\min} \mathbf{1}, \quad (3.33c)$$

$$\mathbf{1}^T \mathbf{x} = 1, \quad (3.33d)$$

$$\mathbf{x} \geq \mathbf{0}. \quad (3.33e)$$

We now have the following lemma leading to Theorem 3. Proof of Lemma 1 appears in the appendix.

**Lemma 1** *If the link gains  $G_{lj}$  are all non-zero, then the LP (3.33) which maximizes the minimum common rate among the links results in all links getting the same rate  $r^*$ , i.e.,  $\mathbf{r}^* = r^* \mathbf{1}$ .*

**Theorem 3** *The solution  $\mathbf{x}^*$  obtained by solving the LP (3.33) is max-min fair.*

The proof of Theorem 3 is immediate from the definition of max-min fairness. The

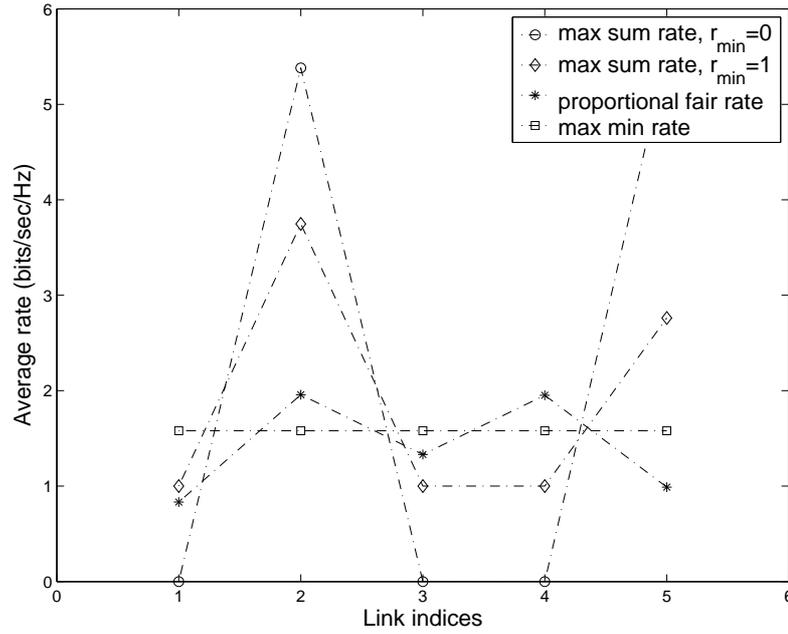


Figure 3.8: Comparison between rates of the links under different settings

above two theorems can be used to show that the solution of the LP (3.33) leads to the max-min fair rates. A similar result appeared in [50] as the *solidarity* property. From Theorem 3, it is clear that solidarity property holds for a Gaussian interference model. Figure 3.7 illustrates the solidarity property for the Gaussian interference model. The solidarity property, however, does not hold for the multiple access model discussed in Section 3.2.

We now find the proportional fair rate vector.

• **Definition:** A vector of rates  $\mathbf{r}$  is *proportional fair* if it is feasible, i.e.,  $\mathbf{C}\mathbf{x} = \mathbf{r}$  for  $\mathbf{x}$  such that  $\mathbf{1}^T \mathbf{x} = 1$  and  $\mathbf{x} \geq \mathbf{0}$ , and if for any other feasible vector  $\mathbf{r}'$ , the aggregate of proportional change is negative.

$$\sum_i \frac{r'_i - r_i}{r_i} \leq 0. \quad (3.34)$$

In [51], Kelly proposed proportional fairness in the context of rate control for elastic traffic. It can be shown that the proportionally fair vector is the one that maximizes the sum of logarithms of the utility functions. Hence, to obtain the proportional fair

rates, we solve the following non-linear optimization problem with linear constraints

$$\max_{\mathbf{r}, \mathbf{x}} \quad \sum_l \log r_l \quad (3.35a)$$

$$\text{subject to} \quad \mathbf{r} = \mathbf{C}\mathbf{x}, \quad (3.35b)$$

$$\mathbf{1}^T \mathbf{x} = 1, \quad (3.35c)$$

$$\mathbf{x} \geq \mathbf{0}. \quad (3.35d)$$

The objective function of the above non-linear optimization problem is increasing and strictly concave. The constraint set is linear and hence the problem is a convex optimization problem [49]. This implies that the problem has a unique global maximum over the constraint set. The solution for such problems can be found out by gradient search algorithms [52]. The comparison of scheduling schemes under different optimization settings is shown in Figure 3.8. Notice that only in the case of maximum sum rate with no minimum rate constraint, there exist links with zero obtained rate. In the case of the max-min fair solution, all the links end up getting the same rate. We observe that proportional fair rates depend on the topology of the links and strike a balance between high rates for some links and reasonable rates for disadvantaged links.

### 3.4.1 Equal time scheduling

In this section, we consider the schedule which provides equal time of activity for each of the links in the network. This may be useful in cases like sensor networks when the nodes have just sufficient energy to transmit some data and all the nodes should be given equal priority in transmission of data. The solution of the following LP gives the equal-time schedule:

$$\max \quad \mathbf{1}^T \mathbf{C}\mathbf{x} \quad (3.36a)$$

$$\text{subject to} \quad \mathbf{T}\mathbf{x} = \tau \mathbf{1}, \quad (3.36b)$$

$$\mathbf{1}^T \mathbf{x} \leq 1, \quad (3.36c)$$

$$\mathbf{x} \geq \mathbf{0}. \quad (3.36d)$$

where,  $\mathbf{T}$  is the binary matrix defined in Section 3.2 and  $\tau$  represents the equal time period of the schedule. Note that whenever the value of  $\tau$  is very small, the transmission mode whose activity vector is all zero is operated (i.e., none of the links are active). When the value of  $\tau$  is very high, the transmission mode whose activity vector is all ones is operated (i.e., all the links are active).

### 3.5 Energy Efficient scheduling

In certain applications, it may be required that the schedule conforms to stringent energy constraints in the network [53]. In this section, we study the energy efficiency of the maximum sum rate schedule. In the sequel, we define energy efficiency formally and modify the LP in Section 3.3 so that the energy efficiency of the schedule is more than a certain threshold.

We first introduce the notion of energy efficiency in scheduling. From the definitions in Section 3.2, the average power  $\bar{P}_l$  used by link  $l$  can be obtained as the time average of the power used by link  $l$  in all the transmission modes. Hence,

$$\bar{P}_l = \sum_i P_l t_{li} x_i. \quad (3.37)$$

We define the *energy efficiency*  $\epsilon_l$  of a link as the ratio of the average rate obtained by link  $l$  to the average power  $\bar{P}_l$  expended by the link. Hence,

$$\epsilon_l = \frac{r_l}{\bar{P}_l}. \quad (3.38)$$

We can view the efficiency of a link as the data rate obtained per unit of power used in transmission on the link. Note that the efficiency depends on the schedule  $\mathbf{x}$  and that  $\epsilon_l$  is undefined if link  $l$  does not transmit during that schedule.

The highest efficiency of a link is attained when it transmits in isolation. We refer to the modes corresponding to a link  $l$  transmitting in isolation without interference from other links as the link  $l$  singleton mode. If the system were to use only the link  $l$

singleton mode, link  $l$  obtains SINR (in fact SNR)

$$\bar{\gamma}_l = \frac{G_l P_l}{\sigma_l^2} \quad (3.39)$$

and efficiency

$$\bar{\epsilon}_l = \frac{\log(1 + \bar{\gamma}_l)}{P_l}. \quad (3.40)$$

For any schedule in which link  $l$  is active, its efficiency  $\epsilon_l$  must satisfy  $\epsilon_l \leq \bar{\epsilon}_l$  since among all the transmission modes, the link obtains its maximum rate in the singleton mode as there are no other interfering users that bring down the rate.

We are interested in obtaining efficient scheduling strategies which maximize the sum rate in the links by ensuring that all links operate above a certain threshold efficiency  $\epsilon_0$ , i.e.,  $\epsilon_l \geq \epsilon_0$  for all links  $l$ . Hence from (3.38),

$$\sum_i c_{li} x_i \geq \epsilon_0 \sum_i P_l t_{li} x_i. \quad (3.41)$$

Thus, equation (3.41) can be written in vector form as

$$\mathbf{C}\mathbf{x} \geq \epsilon_0 \mathbf{P}\mathbf{T}\mathbf{x}. \quad (3.42)$$

The energy efficient scheduling problem can be presented as the LP:

$$\max_{\mathbf{r}, \mathbf{x}} \quad \mathbf{1}^T \mathbf{r} \quad (3.43a)$$

$$\text{subject to} \quad \mathbf{r} = \mathbf{C}\mathbf{x}, \quad (3.43b)$$

$$\mathbf{r} \geq r_{\min} \mathbf{1}, \quad (3.43c)$$

$$\mathbf{r} \geq \epsilon_0 \mathbf{P}\mathbf{T}\mathbf{x}, \quad (3.43d)$$

$$\mathbf{1}^T \mathbf{x} \leq 1, \quad (3.43e)$$

$$\mathbf{x} \geq \mathbf{0}. \quad (3.43f)$$

where (3.43c) represents the minimum rate constraint and (3.43d) is the efficiency

constraint. Rewriting the LP (3.43) in terms of the variable  $\mathbf{x}$  only, we get

$$c_{\text{opt}}(r_{\min}, \epsilon_0) = \max_{\mathbf{x}} \quad \mathbf{1}^T \mathbf{C} \mathbf{x} \quad (3.44a)$$

$$\text{subject to} \quad \mathbf{C} \mathbf{x} \geq r_{\min} \mathbf{1}, \quad (3.44b)$$

$$(\mathbf{C} - \epsilon_0 \mathbf{P} \mathbf{T}) \mathbf{x} \geq \mathbf{0}, \quad (3.44c)$$

$$\mathbf{1}^T \mathbf{x} \leq 1, \quad (3.44d)$$

$$\mathbf{x} \geq \mathbf{0}. \quad (3.44e)$$

We now consider the special case  $r_{\min} = 0$  when there is no minimum rate requirement for any of the links. Before proceeding, we observe that  $c_{li} = 0$  if and only if  $t_{li} = 0$ . Thus for sufficiently small efficiency threshold  $\epsilon_0$ ,  $\mathbf{C} - \epsilon_0 \mathbf{P} \mathbf{T}$  is a nonnegative matrix. In this case, the constraint (3.44c) is inactive in that it is trivially satisfied for all nonnegative  $\mathbf{x}$ .

In the absence of the efficiency constraint, as discussed in Section 3.3, the optimal schedule for the problem (3.10) with  $r_{\min} = 0$  is to operate only a transmission mode  $d$  with maximum sum rate. The optimal objective value is the maximum column sum of the rate matrix  $\mathbf{C}$ . We refer to mode  $d$  as the dominant mode and we denote by  $\hat{\mathbf{x}} = [0 \cdots 0 \ 1 \ 0 \cdots 0]^T$  the schedule that supports exclusive use of mode  $d$ . In addition, we use  $\mathcal{D}$  to denote the set of active links in the dominant mode. For links  $l \in \mathcal{D}$ , we use  $\hat{\epsilon}_l$  to denote the efficiency of the link under the schedule  $\hat{\mathbf{x}}$ .

When the efficiency threshold  $\epsilon_0$  satisfies

$$\epsilon_0 \leq \hat{\epsilon} = \min_{l \in \mathcal{D}} \hat{\epsilon}_l, \quad (3.45)$$

all links in the dominant mode will meet the efficiency constraint under schedule  $\hat{\mathbf{x}}$ . In this case, schedule  $\hat{\mathbf{x}}$ , corresponding to exclusive use of the dominant mode, remains optimal. When the efficiency threshold  $\epsilon_0$  passes  $\hat{\epsilon}$ , the optimal schedule may continue to employ the dominant mode, but other modes also must be scheduled to boost the efficiency of the least efficient links in the dominant mode. Eventually, the scheduling becomes infeasible when  $\epsilon_0$  exceeds  $\bar{\epsilon} = \max_l \bar{\epsilon}_l$ , the maximum efficiency in a singleton

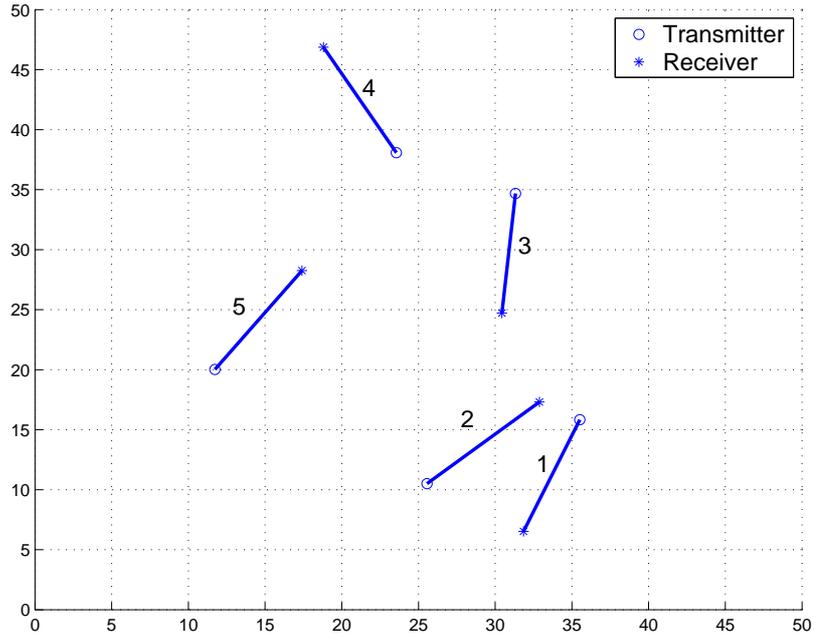


Figure 3.9: Set of five links each of length  $d = 10$ .

mode.

Depending on the geometry of the links, the dominant transmission mode can be a single active link or a collection of geographically separated links. However, the geographic separation of links that is typically associated with the dominant mode is consistent with those links in the dominant mode having high efficiency. The consequence is that enforcing an efficiency constraint typically has little impact until the efficiency constraint is very stringent.

When we impose a non-zero minimum rate  $r_{\min}$  on all links, the optimal schedule may change. Once again, for small values of the threshold  $\epsilon_0$ , the efficiency constraint is inactive. Increasing the threshold  $\epsilon_0$  gradually eliminates modes in which links are active in the presence of significant interference. For high values of  $\epsilon_0$ , the system tends to use the high-efficiency singleton modes, although this can have a significant penalty in terms of the sum rate.

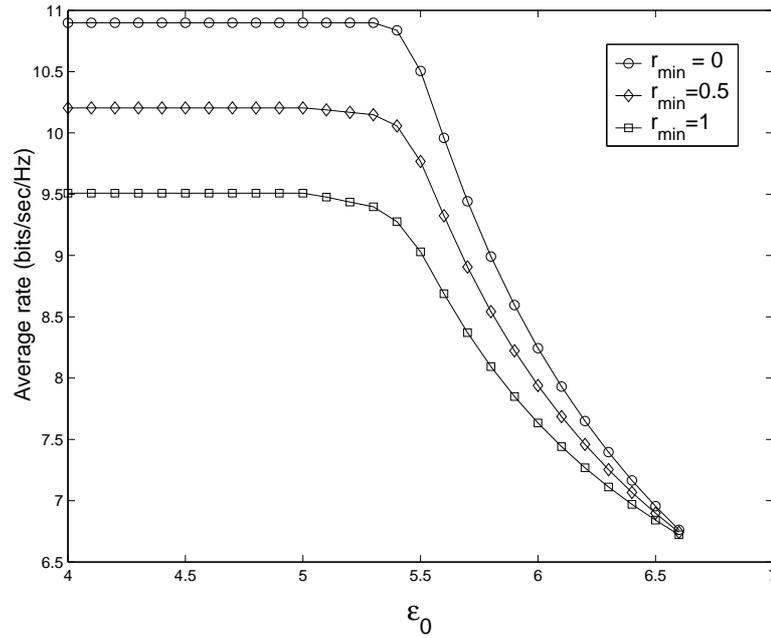


Figure 3.10: Variation of sum rate with efficiency  $\epsilon$

### 3.5.1 Simulation Results

We now illustrate energy efficient scheduling using our example network shown in Figure 3.9. In the case of maximum sum rate scheduling with no minimum rate constraint, for a fixed  $\epsilon_0 = \epsilon$ , the transmission mode with the highest sum rate is chosen. The links which are not a part of the dominant transmission mode are not operated at all. As we increase  $\epsilon$ , the sum rate remains constant until a certain threshold value, say  $\epsilon_{th}$ . The value of  $\epsilon_{th}$  is the efficiency of the weakest (in terms of link quality) link in the dominant mode. For  $\epsilon > \epsilon_{th}$ , single link modes corresponding to some of the links in the dominant mode are operated in order to satisfy the efficiency constraints on those links. The variation of sum-rate with  $\epsilon$  is shown in Figure 3.10.

In the case of maximum sum rate scheduling with non-zero minimum rate constraint, we see that more than one transmission mode is operated since there is a minimum rate requirement for each link. As with the case with zero minimum rate constraint, for a given degree of fairness (i.e., a specified  $r_{\min}$ ), there is relatively little penalty for requiring efficiency, until the breakpoint where the required efficiency approaches the

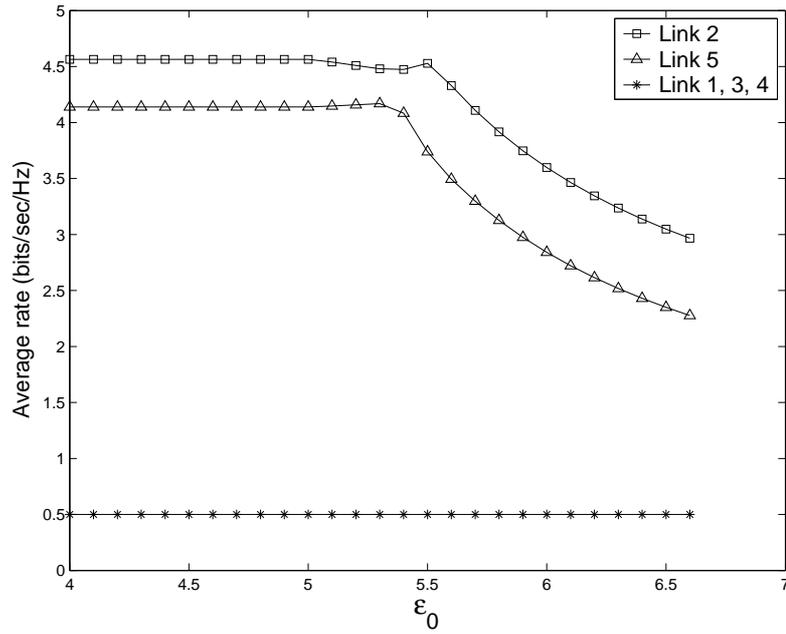


Figure 3.11: Variation of rates of individual links with efficiency  $\epsilon$  for  $r_{\min} = 0.5$

efficiency of the link in isolation. For larger values of  $r_{\min}$ , the efficiency of the single link modes decreases. This is illustrated in Figure 3.10.

Notice that for non-zero  $r_{\min}$  values, there is some loss in sum rate because of the fairness introduced by  $r_{\min}$ . However, similar to the case when  $r_{\min} = 0$ , when  $\epsilon_0$  crosses a threshold, the penalty is sum rate increases as efficiency increases.

Figure 3.11 shows the variation of rates of individual links with the efficiency. For lower values of  $\epsilon$ , the dominant mode  $\{1, 4\}$  is operated to maximize the sum rate and the links which are not a part of the dominant mode are scheduled for just enough time to satisfy their minimum rate constraint. When higher efficiency is required, the rates of links in the dominant mode reduces until the singleton modes are not efficient any more.

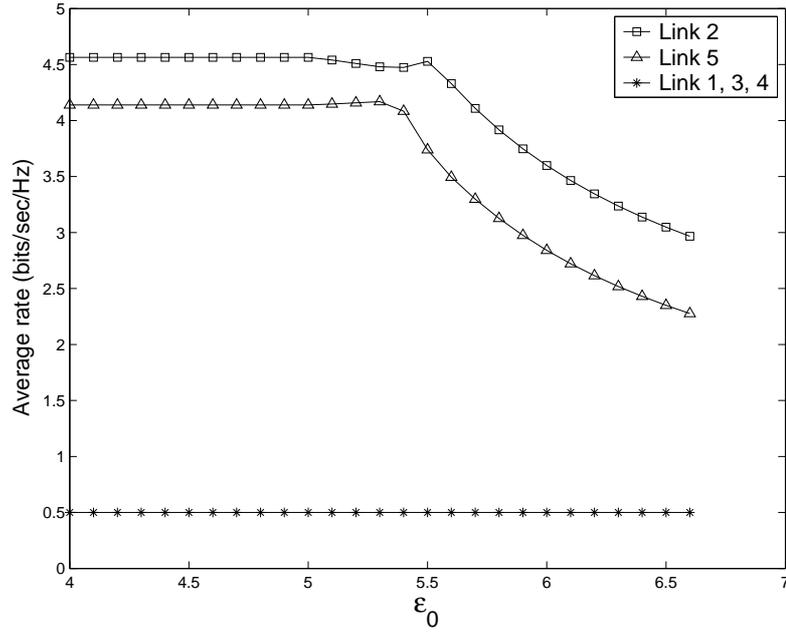


Figure 3.12: Variation of rates of individual links with efficiency  $\epsilon$  for  $\mathbf{r}_{\min} = 0.5$

### 3.6 A unified framework for centralized scheduling

There may be cases in a heterogenous network where the nodes of the network have various rate and efficiency requirements. In this section, we present the unified optimization problem for the spectrum server which obtains the desired schedule when some links have minimum rate requirements and some of the links come with a fixed energy efficiency constraint. Let us define the matrix

$$\mathbf{E} = \text{diag}(\epsilon_1, \epsilon_2, \dots, \epsilon_L). \quad (3.46)$$

Note that for those links for which  $\epsilon_l = 0$ , there is no constraint on the energy efficiency. We now present the optimization problem for the case when there are heterogenous

requirements for the nodes in the network.

$$\max \quad f(\mathbf{r}) \quad (3.47a)$$

$$\text{subject to} \quad \mathbf{r} = \mathbf{C}\mathbf{x}, \quad (3.47b)$$

$$\mathbf{r} \geq \mathbf{r}_{\min}, \quad (3.47c)$$

$$\mathbf{r} \geq \mathbf{EPT}\mathbf{x}, \quad (3.47d)$$

$$\mathbf{x} \in \mathcal{X}. \quad (3.47e)$$

where in (3.47c), the rate vector  $\mathbf{r}_{\min}$  may contain zeroes to reflect the absence of a strict rate requirement for specific users and in (3.47d), the energy efficiency constraint of each link is captured. If  $\mathbf{E}$  is a zero matrix, then above problem (3.47) reduces to the maximum sum rate schedule and if  $\mathbf{r}_{\min} = \mathbf{0}$ , then the problem reduces to energy efficient scheduling discussed in Section 3.5. The objective function  $f(\mathbf{r})$  is assumed to be a concave, non-decreasing function of the rate vector  $\mathbf{r}$ . For example, if  $f(\mathbf{r}) = \min_i r_i$ , then the solution to (3.47) is the maximum common rate schedule. If  $f(\mathbf{r}) = \sum_i \log(r_i)$ , then (3.47) gives the proportional fair rates. Figure 3.12 shows the variation of the individual rates in the links with efficiency when the minimum rate requirement is different. In this particular setting, there is no efficiency constraint on the links 2 and 5.

### 3.7 Towards a distributed algorithm - Random scheduling

The centralized scheduling strategies that have been presented thus far, require information about all the links to be available at the spectrum server. There are many reasons why such a centralized scheduler may be difficult to implement in a real world situation, including:

1. Complete information about all the links and their channel gains should be known to the spectrum server to solve the scheduling problem precisely.
2. If the number of the links increases, the size of the LP increases exponentially.

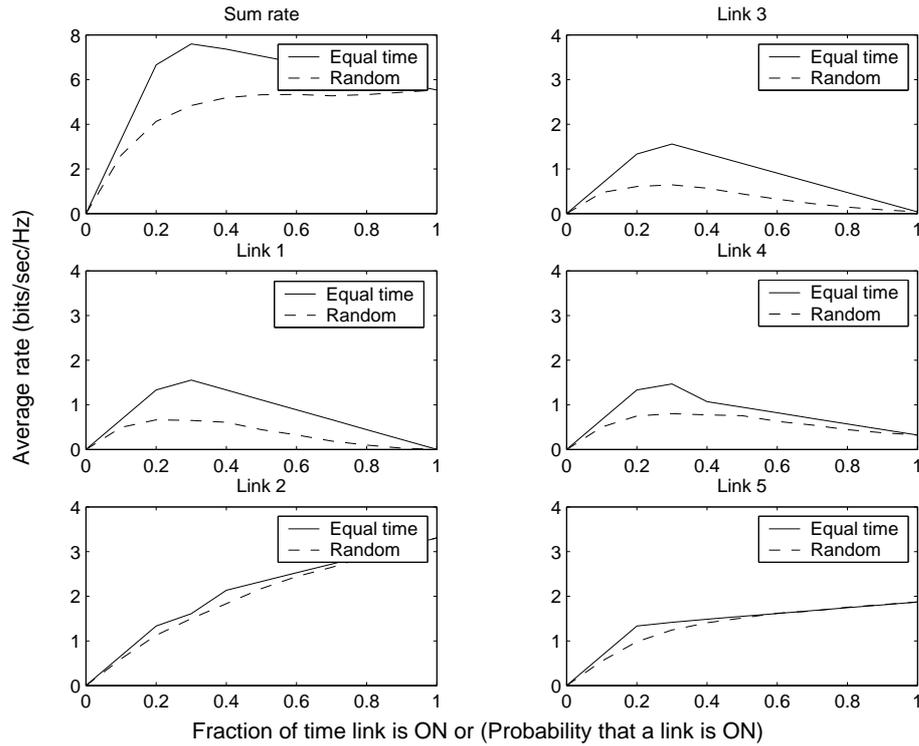


Figure 3.13: Comparison of the average rates obtained by equal time scheduling and random probabilistic scheduling.

3. The exchange of information between the centralized spectrum server and the individual transmitters may not be an easy task.

Hence, a distributed algorithm to implement these scheduling schemes would be a desirable solution. Towards that end, we consider a simple distributed scheduling mechanism called the random scheduling scheme.

In this scheme, time is slotted and each link  $l$  turns itself on or off based on a fixed probability  $p_l$ , independent of the other links. Depending on the interference from the other links, the instantaneous rates in all links are obtained. This simple on-off scheme is completely distributed in the sense that the interference gains between any two links need not be known to any other link. Though this random scheduling scheme is easy to implement, to get better sum rates, it is required to optimize over the probabilities of the activity of each link. It may not be possible to obtain all possible rates that are attained by the centralized scheme.

In Figure 3.13, we compare the rates obtained by the random scheduling and equal time scheduling strategy. We see that the rates of some of the individual links in the random scheduling scheme are close to those obtained by the equal time scheduling scheme, especially in the dominant mode. This motivates the need to look for distributed scheduling algorithms that can achieve the various fairness criteria discussed here in the context of centralized scheduling. In [54], we compare the throughput regions of centralized scheduling and a probabilistic random access scheme, wherein in each slot, a link is active with a fixed probability chosen independent of other interfering links. We observe that for the case of two interfering links, the probabilistic scheme does not suffer any loss in the rate region relative to the centralized scheme if the interference between the links is sufficiently low. For more than two interfering links, the characterization of throughput rate region for the probabilistic scheme becomes intractable and similar observations are not easily forthcoming. However, we give a distributed algorithm where each link independently updates its transmission probability based on its measured throughput to achieve any desired feasible rate vector in the throughput region of the probabilistic scheme and prove its convergence.

### 3.8 Conclusion

We introduced the notion of a spectrum server, which allocates a schedule for a set of links in a wireless network, which is modeled as a directed graph. The problem of maximizing the sum rate in all the links subject to minimum rate constraints and energy efficiency constraints was posed as a linear program. With knowledge of the link gains in the network, the spectrum server scheduled the on/off periods of the links so as to maximize the sum rate and/or satisfy constraints on link fairness and efficiency. We provided two classes of schedules — one which maximizes the sum rate subject to minimum rate constraints in the link and the other schedule which maximized the sum rate subject to fairness and efficiency constraints. We also derived the schedule that maximized the common rate. We introduced equal time scheduling as a special case of fair scheduling. In the special case when there was no minimum rate constraint, varying the efficiency constraint caused the optimal policy to vary from a fixed dominant

mode with highest sum rate being operated all the time to time sharing among singleton modes in which just one link is active.

### 3.9 Appendix

*Proof of Lemma 1:*

Let  $r^*$  be the optimal value of LP (3.33) which maximizes the common minimum rate. Let  $\mathbf{x}^*$  be the optimal schedule corresponding to the of active transmission modes  $\mathcal{T}^* = \{i \in \mathcal{T} : x_i^* > 0\}$ . Note that the idle transmission mode with the all zero activity vector would never be a part of  $\mathcal{T}^*$  because, if it were, we can improve the rates of links in  $\mathcal{L}_2$  and this contradicts that  $r^*$  is the optimal solution of (3.33). It is required to prove that at optima, the rate vector  $\mathbf{C}\mathbf{x} = r^*\mathbf{1}$ . We assume the contrary that the solution to (3.33) leads to unequal rates over the set of  $L$  links. We can then partition the sets of links  $\mathcal{E}$  into two disjoint non-empty sets  $\mathcal{L}_1 = \{l \in \mathcal{E} : r_l > r^*\}$  and  $\mathcal{L}_2 = \{l \in \mathcal{E} : r_l = r^*\}$ . This in turn induces a partition on the set  $\mathcal{T}^*$  of all active transmission modes for the optimal solution into three disjoint sets  $\mathcal{T}_1^*$ ,  $\mathcal{T}_2^*$  and  $\mathcal{T}_3^*$  such that

$$\mathcal{T}_1^* = \{i \in \mathcal{T}^* : t_{il} = 0, \text{ for all } l \in \mathcal{L}_2\}, \quad (3.48)$$

$$\mathcal{T}_2^* = \{i \in \mathcal{T}^* : t_{il} = 0, \text{ for all } l \in \mathcal{L}_1\}, \quad (3.49)$$

$$\mathcal{T}_3^* = \mathcal{T}^* \setminus \{\mathcal{T}_1^* \cup \mathcal{T}_2^*\}. \quad (3.50)$$

$\mathcal{T}_1^*$  and  $\mathcal{T}_2^*$  contain active transmission modes which consist of links only from  $\mathcal{L}_1$  and  $\mathcal{L}_2$  respectively, and  $\mathcal{T}_3^*$  contains transmission modes which consist of links in both  $\mathcal{L}_1$  and  $\mathcal{L}_2$ . We consider two cases below.

### 3.9.1 Case (i): $\mathcal{T}_1^*$ is non-empty

There exists an active transmission mode  $i \in \mathcal{T}_1^*$  consisting of links only from  $\mathcal{L}_1$ . Consider the mode  $i'$  with activity vector  $t_{i'}$  given by

$$t_{i'} = \begin{cases} 1, & \text{for all } l \in \mathcal{L}_2, \\ 0, & \text{otherwise.} \end{cases} \quad (3.51)$$

All the links from the set  $\mathcal{L}_2$  are active under mode  $i'$ . Therefore,  $c_{li'} > 0$  for  $l \in \mathcal{L}_2$ . In the optimal schedule  $\mathbf{x}^*$ , we know that  $x_i^* > 0$  but  $x_{i'}^*$  may be zero. The rate in link  $l$  under schedule  $\mathbf{x}^*$  is  $r_l = \sum_k c_{lk} x_k^*$ . Define for some fixed  $\epsilon_1 > 0$ , the schedule

$$\hat{\mathbf{x}} = [x_1^* \dots x_i^* - \epsilon_1 \dots x_{i'}^* + \epsilon_1 \dots x_{M-1}^*]^T. \quad (3.52)$$

For sufficiently small  $\epsilon_1$ , the schedule  $\hat{\mathbf{x}}$  will be feasible. Now, for  $l \in \mathcal{L}_2$ , the rate  $\hat{r}_l$  due to schedule  $\hat{\mathbf{x}}$  is

$$\hat{r}_l = \sum_k c_{lk} \hat{x}_k = r^* - c_{li} \epsilon_1 + c_{li'} \epsilon_1 \quad (3.53)$$

Since  $c_{li'} > 0$  and  $c_{li} = 0$  for  $l \in \mathcal{L}_2$ ,

$$\hat{r}_l = r^* + c_{li'} \epsilon_1. \quad (3.54)$$

Thus, we conclude that  $\hat{r}_l > r^*$ ,  $l \in \mathcal{L}_2$ . Note that  $\epsilon_1$  needs to be chosen such that for all  $l \in \mathcal{L}_1$ ,  $\hat{r}_l > r^*$ . The choice of  $\epsilon_1$  such that  $c_{li} \epsilon_1 < \min_{l \in \mathcal{L}_1} r_l - r^*$  ensures that  $\hat{r}_l > r^*$  for  $l \in \mathcal{L}_1$ . We can thus improve the rates in all links in  $\mathcal{L}_2$ . This contradicts the optimality of  $r^*$ . We denote this step as *Increase(1)*.

### 3.9.2 Case (ii): $\mathcal{T}_1^*$ is empty

In this case, if  $\mathcal{T}_3^*$  is empty, then  $\mathcal{T}^* = \mathcal{T}_2^*$  and hence all rates are equal, and the proof is complete. Thus we consider only the case of  $\mathcal{T}_3^*$  being non-empty. For an active mode  $j \in \mathcal{T}_3^*$ , there exist non-empty subsets of  $\mathcal{L}_1$  and  $\mathcal{L}_2$ , namely  $\mathcal{M}_1$  and  $\mathcal{M}_2$  such that the

activity vector  $t_j$  is given by

$$t_{lj} = \begin{cases} 1, & l \in \mathcal{M}_1 \subseteq \mathcal{L}_1, \\ 1, & l \in \mathcal{M}_2 \subseteq \mathcal{L}_2, \\ 0, & \text{otherwise.} \end{cases} \quad (3.55)$$

Consider the mode  $j'$  for which the activity vector  $t_{j'}$  is given by

$$t_{lj'} = \begin{cases} 1, & \text{if } l \in \mathcal{M}_2, \\ 0, & \text{otherwise.} \end{cases} \quad (3.56)$$

We have assumed that all link gains  $G_{lj}$  are non-zero, that is there is lesser interference for links in  $\mathcal{M}_2$  in mode  $j'$  than in mode  $j$  due to a lesser number of active links in mode  $j'$ . Thus for links  $l \in \mathcal{M}_2$ ,

$$c_{lj} = \frac{G_{lj}P_l}{\sum_{k \in \mathcal{M}_1 \cup \mathcal{M}_2, k \neq l} t_{kj}G_{lk}P_k + \sigma_l^2} < \frac{G_{lj}P_l}{\sum_{k \in \mathcal{M}_2, k \neq l} t_{kj'}G_{lk}P_k + \sigma_l^2} = c_{lj'}. \quad (3.57)$$

Since  $j \in \mathcal{T}_3^*$ ,  $x_j^* > 0$ . For some  $\epsilon_2 > 0$ , we define a feasible schedule

$$\hat{\mathbf{x}} = [x_1^* \dots x_j^* - \epsilon_2 \dots x_{j'}^* + \epsilon_2 \dots x_{M-1}^*]^T. \quad (3.58)$$

Under schedule  $\mathbf{x}^*$  and  $\hat{\mathbf{x}}$ , link  $l$  obtains rate  $r_l = \sum_k c_{lk}x_k^*$  and  $\hat{r}_l = \sum_k c_{lk}\hat{x}_k$  respectively. Thus, the difference

$$\hat{r}_l - r_l = (\hat{x}_j - x_j^*)c_{lj'} + (\hat{x}_j - x_j^*)c_{lj} = \epsilon_2(c_{lj'} - c_{lj}). \quad (3.59)$$

It follows from (3.57) that  $\hat{r}_l - r_l > 0$  for  $l \in \mathcal{M}_2$ . Let us call this step *Increase(2)*.

Since  $\mathcal{L}_2$  is a finite set, repeatedly applying *Increase(1)* or *Increase(2)* on  $\mathcal{L}_2 \setminus \mathcal{M}_2$ , we can increase the rates of all the links in  $\mathcal{L}_2$ . This contradicts the optimality of  $r^*$ . The proof is complete since both cases contradict the fact that the optimal solution leads to unequal rates in the links. ■

## Chapter 4

### Cross Layer Scheduling in Multihop Networks via a Spectrum Server

In this chapter, we extend the framework described in the previous chapter to the case where the message from a node has to traverse through multiple nodes in the network. Traversing through multiple nodes in the wireless network may be a desirable feature since the source may transmit with lesser power to the neighboring node, thus relaying the message via other nodes in the network. This reduces the total interference in the network and allows for other simultaneously reliable transmissions in the network.

#### 4.1 Cross-layer Resource Allocation of End-to-End Flows

Figure 4.1 shows a multihop wireless network transporting data from the source (node 1) to the destination (node 6) in a series of hops. In general, a network could have many simultaneously active sessions, each specified by an origin-destination pair. Each session may split its traffic into multiple flows through distinct routes in the network. For instance, Figure 4.1 shows the session from node 1 to node 6 split into two flows  $f_1$  and  $f_2$  through the network. The sessions may individually demand QoS requirements such as minimum average rate, maximum peak rate, limited delay, bounded jitter etc. In order to satisfy the QoS guarantees, each layer in the protocol stack optimizes a set of utility functions in view of achieving the global objectives (QoS agreements) across the network. For instance, the network layer specifies the appropriate routes for the individual flows in the network. The decision to choose a particular route may be based on some metric chosen according to the QoS requirements of each session. In most cases, the routing metric is a function of the service delays and the average rates supported by the links. Moreover, the routing decisions may have to be updated in

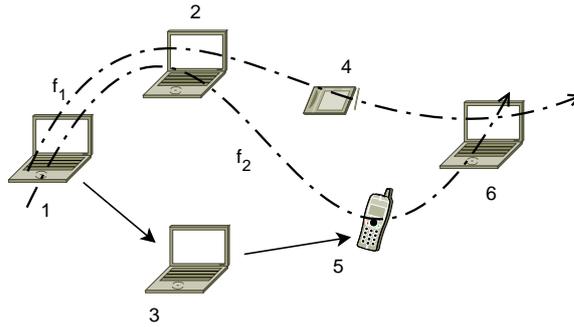


Figure 4.1: An example of a simple multihop network. The solid arrows are instances of links in the network and dotted lines show a flow through the network

regular intervals of time to account for the variation of the parameters in underlying layers.

The MAC/PHY layers solve the resource allocation problem. These layers specify the rate and power allocation, transmission strategy and schedule for links in the network. The *power allocation* strategy on a link depends on the channel conditions (its knowledge or lack thereof at each transmitter node), constraints on the transmission power due to device limitations and the overall interference of other transmitters in its neighborhood. The *instantaneous rate* obtained in a link depends on the underlying channel propagation parameters, modulation and coding scheme used at the transmitter, signal processing employed at the receiver and the interference in the neighborhood of the receiver. In general, a MAC scheme specifies a set of rules for transmissions in the network. The choice of rules depend on the amount of resources allocated to the links in the network. Addressing the resource allocation problem in the MAC/PHY layer taking into account all these variables can be mathematically intractable [55,56].

Many individual aspects in different layers of a multihop network have been addressed by the research community by making simplifying assumptions on the rest of the network. In particular, there have been numerous advances in the physical layer technologies over the past two decades — adaptive modulation, reliable coding schemes (e.g., LDPC codes, turbo codes etc.), hybrid ARQ schemes, OFDM, MIMO processing, cooperative diversity schemes, multiple access techniques, multiuser diversity schemes to name a few. However, most of the cross-layer design approaches addressed in the

literature make simple and specific assumptions on the underlying physical layer technology. As a result, the impact of these advanced physical layer techniques is not very well understood in a practical setup of a multihop network supporting end-to-end flows. To the best of our knowledge, a unified cross-layer design framework encompassing all physical layer technologies has not previously been proposed.

#### 4.1.1 Related Work

Optimization-based cross-layer resource allocation in wireless networks has been an active area of research and has been studied in different settings (see [48] and the references therein). Since the link rates in a wireless network may vary depending on the various power allocation policies, the set of feasible link rates is a non-convex set. Hence the scheduling problem is typically a non-convex problem. Thus, providing analytical solutions to such cross-layer problems is often difficult [46]. In [57], the authors pose a problem to optimize the source data rates so that they match the link rates to maintain stability of all queues in the network. The solution decomposes to a congestion control component and a scheduling component, with the queue length information as implicit costs. Although the congestion control problem can be solved in a distributed fashion, the scheduling problem is a global optimization problem of high complexity. Some authors [41, 43] consider power control as against variable rate transmission while jointly optimizing routing and scheduling decision across layers. Other works [58–60] consider the protocol model [37] in which the neighboring interferers are not allowed to transmit and each transmission yields a fixed rate in the link. These models correspond to strict layering of the protocol stack and such models are easily amenable to results from fixed wireline networks. In [61], the authors present the SRRAS (simultaneously optimal routing, resource allocation and scheduling) problem and propose a nonlinear column generation technique to reduce the complexity of the scheduling problem. We present a model similar to [61], with an emphasis on unification of a wide variety of physical layer models and understand the effect of some physical layer techniques, e.g., successive decoding techniques on cross-layer routing and scheduling.

This paper extends our earlier work [62, 63] to a centralized cross-layer optimization

framework in which we separate the functionalities of network layer and the MAC/PHY layers. Studies employing similar models can be found in [43, 61, 64, 65]. We describe the relationship of our work with [64] in Section 4.3.3. In our work, we assume that the network layer specifies the set of all feasible routes in the network. Together with the instantaneous rates that can be attained in the network for different configurations of active links, the spectrum server computes a set of flows and a series of feasible schedules of link transmissions that make up the routes for the flows. A key feature of the framework is a physical layer model that constitutes an enumeration of a set of instantaneous attainable rates each for a given transmission mode of the system. We optimize the schedule of transmission modes in the network to maximize a utility function that may include throughput and fairness objectives. Special cases of the model include in-band interference between the links, successive interference cancellation at the link receivers, and multiple antennas at the link transmitters and receivers.

## 4.2 System Model

In this section, we present the physical layer model followed by the network layer flow model. We start with a summary of the notation used in this paper. We use boldface lowercase characters for vectors and boldface uppercase for matrices. If  $\mathbf{a}$  is a vector,  $\mathbf{a}^T$  denotes its transpose and  $\mathbf{a}^T \mathbf{b} = \sum_i a_i b_i$  represents the inner product of the vectors  $\mathbf{a}$  and  $\mathbf{b}$ . We denote the  $(i, j)$ th element of a matrix  $\mathbf{X}$  as  $[\mathbf{X}]_{ij}$ . Consider a wireless network with  $N$  nodes forming  $L$  links sharing a common spectrum. The network can be represented as a directed graph  $\mathcal{G}(\mathcal{V}, \mathcal{E})$ , where the nodes in the network are represented by the set of vertices  $\mathcal{V}$  of the graph and the logical links are represented by directed set of edges  $\mathcal{E}$ .

### 4.2.1 Physical Layer and Interference Model

We use the physical layer model with transmission modes described in the previous chapter. The model has the following two benefits:

1. The model is quite general and it covers various physical layer transmission and

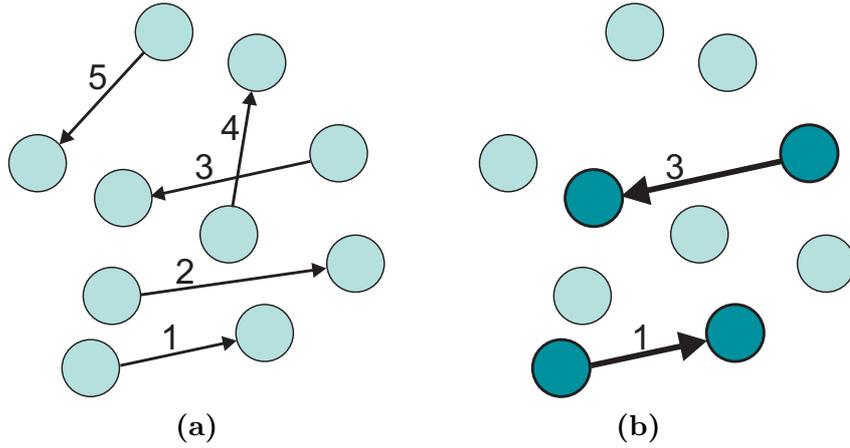


Figure 4.2: Graph of a network: (a) Nodes and the directed links (b) Illustration of transmission mode  $\mathbf{t}=[1 \ 0 \ 1 \ 0 \ 0]$  in which links 1 and 3 are active

signal processing schemes, and

2. The model captures all the necessary information, that the higher layers can make use of to base their optimization decisions upon.

Recall that  $\mathcal{T} = \{0, 1, \dots, M - 1\}$  denotes the set of transmission modes of a network, where  $M$  denotes the number of possible transmission modes. Then the *mode activity vector*  $\mathbf{t}_k$  of mode  $k \in \mathcal{T}$  is a binary vector, indicating the on-off activity of the links. If  $\mathbf{t}_k = (t_{1k}, t_{2k}, \dots, t_{Lk})$  is a mode activity vector, then

$$t_{lk} = \begin{cases} 1, & \text{link } l \text{ is active under transmission mode } k, \\ 0, & \text{otherwise.} \end{cases} \quad (4.1)$$

Note that not all  $M$  modes may be valid for transmission depending on the constraints on the system. For instance, since the links share a common spectrum, the transmitter and receiver in a node operate in the same channel. Hence the node cannot transmit and receive simultaneously because of self-interference. Figure 4.2 shows a representative network and a particular transmission mode for the set of links.

We also recall the rate vector in links,

$$\mathbf{r} = \mathbf{C}\mathbf{x}, \quad (4.2)$$

where  $\mathbf{C}$  is a  $L \times M$  matrix with non-negative entries  $[\mathbf{C}]_{li} = c_{li}$ ,  $\mathbf{r}$  is a real vector of length  $L$  and  $\mathbf{x}$  is a real vector of length  $M$ . Embedded in the matrix  $\mathbf{C}$  are the rates obtained in each link  $l$  as a part of transmission mode with simultaneous transmissions on multiple links.

For our analytical model, we optimize these average rates per frame by specification of the time fractions in  $\mathbf{x}$ , without explicitly specifying the precise slots assigned to each mode. We denote the set of all feasible schedule vectors by

$$\mathcal{X} = \{\mathbf{x} : \mathbf{1}^T \mathbf{x} = 1, \mathbf{x} \geq \mathbf{0}\}. \quad (4.3)$$

We note that this model allows for consideration of a large class of physical layer interactions. We observe that all necessary aspects of transmitter and receiver technology are embedded in the rate matrix  $\mathbf{C}$ .

#### 4.2.2 Traffic flow model

The traffic model we consider is similar to the multi-commodity flow model discussed in [51]. We work with average rates on the links and assume a fluid flow model for the routing problem. We also assume that each node in the network has infinite buffer and sources have an infinite backlog of data to send through the network. We assume that the network consists of  $K$  sessions. A *session* is the end-to-end data originating at a source node in the network and ending at a destination node. A session may be split into multiple flows that travel through different routes in the network. A *route*  $r$  is a sequence of links forming a path in the graph  $\mathcal{G}$  of the network. Let there be  $R$  possible routes in the whole network. We assume that the routes are known at the network layer. The routes in the network are specified by the  $L \times R$  matrix  $\mathbf{A}$  with entries  $\in \{0, 1\}$ , where

$$[\mathbf{A}]_{lr} = \begin{cases} 1, & \text{if link } l \text{ is a part of route } r, \\ 0, & \text{otherwise.} \end{cases} \quad (4.4)$$

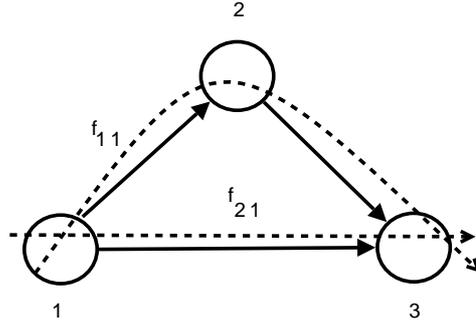


Figure 4.3: Graph of network showing the nodes and directed links. The session is split into two flows  $f_{11}$  and  $f_{21}$ .  $f_{11}$  flows through links (1,2) and (2,3) while  $f_{21}$  goes through (1,3).

Let  $f_{jk}$ ,  $j = 1, \dots, R$ ,  $k = 1, \dots, K$  be the flow corresponding to the  $k$ th session in the  $j$ th route. If  $\mathcal{R}_l$  denotes the set of routes passing through the link  $l$ , we can write the expression for the rate in link  $l$  in the session  $k$  as

$$r_{lk} = \sum_{j \in \mathcal{R}_l} f_{jk} = \mathbf{a}_l^T \mathbf{f}_k, \quad (4.5)$$

where  $\mathbf{f}_k = [f_{1k} \ f_{2k} \ \dots \ f_{Rk}]^T$  is the vector of flows for the  $k$ th session,  $k = 1, \dots, N$ , and  $\mathbf{a}_l^T$  is the  $l$ th row of  $\mathbf{A}$ . Let  $\mathbf{r}_k = [r_{1k} \ r_{2k} \ \dots \ r_{Lk}]^T$ , we can then write the link rate vector equation from (4.5) as,

$$\mathbf{r}_k = \mathbf{A} \mathbf{f}_k. \quad (4.6)$$

Thus the aggregate rates through links  $l = 1, 2, \dots, L$  are given by

$$\mathbf{r} = \sum_k \mathbf{A} \mathbf{f}_k. \quad (4.7)$$

Let us consider an example in Figure 4.3. There are  $L = 3$  links in the network. Consider a single session ( $K = 1$ ) originating at 1 and ending at 3. There are two routes in the network for this session, i.e.,  $R = 2$ . Let the session be split into two flows of rates  $f_{11}$  and  $f_{21}$  along the two routes. If  $\mathbf{f}_1 = [f_{11} \ f_{21}]^T$  is the flow vector for this

session, then the rate vector in the links is given by

$$\mathbf{r}_1 = \begin{bmatrix} r_{11} \\ r_{21} \\ r_{31} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} f_{11} \\ f_{21} \end{bmatrix}. \quad (4.8)$$

The sum of the flows  $f_{11} + f_{21}$  gives the total end-to-end session rate.

Given the routes in the network, we will be interested in maximizing  $\sum_k U_k(\mathbf{f}_k)$ , the sum of utility functions of the rates in each session, where  $U_k : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  is a non-negative, concave, non-decreasing function that is twice differentiable. The optimization problem for maximizing the sum of utility functions of the rates in each session can be posed as the mathematical program:

$$\max_{\mathbf{x}, \mathbf{f}_k} \quad \sum_k U_k(\mathbf{f}_k) \quad (4.9a)$$

$$\text{subject to} \quad \mathbf{r} = \mathbf{C}\mathbf{x}, \quad (4.9b)$$

$$\mathbf{r} \geq \sum_k \mathbf{A}\mathbf{f}_k, \quad (4.9c)$$

$$\mathbf{x} \in \mathcal{X}, \quad (4.9d)$$

$$\mathbf{f}_k \geq 0, \quad k = 1, \dots, K. \quad (4.9e)$$

In the above LP, some of the entries in the vector  $\mathbf{f}_k$  are zero since some of the flows are not part of some routes. If the utility function  $U_k(\mathbf{f}_k) = \sum_j \alpha_{kj} f_{jk}$ , where  $\alpha_{kj}$  represent the weights for flows, then (4.9) maximizes a weighted sum of end-to-end flows of sessions. We then get a linear program that can be solved using standard techniques [66]. If  $U_k(\mathbf{f}_k) = w_k \log(\mathbf{1}^T \mathbf{f}_k)$ , then (4.9) solves for the weighted proportional fair rates of the flows. More general forms of utility functions for elastic traffic corresponding to various fairness objectives are given in [67].

The optimization problem described in (4.9) is a cross-layer optimization problem where, the average rates in the links at the MAC/PHY layers are specified by time schedules. Constraints (4.9b) and (4.9c) imply that the sum of the flows in each link  $l$  is upper bounded by an average rate  $r_l$  given by the link schedule determined by the

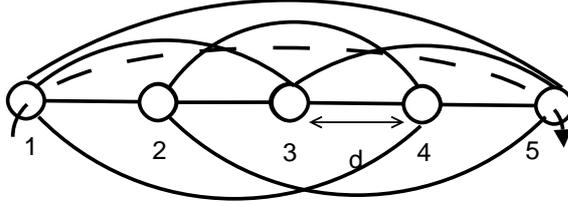


Figure 4.4: Linear network of 5 nodes, with inter-node distance  $d$ . The network has 10 links. A session originating at node 1 and terminating at node 5 is shown. The session can take 8 distinct routes in the network.

MAC/PHY layers. The spectrum server solves the cross-layer routing and scheduling problem, specifying the amount of flow carried by each route to match the average link rates. This is in contrast with the model described in [51], each link has a fixed finite capacity and this problem is solved at the network layer only. The result of the optimization program (4.9) is a set of transmission modes along with the time fraction of operation of these modes (solution to the scheduling problem) and the flows in each route. Appropriate activity of the modes makes the transport of end-to-end flows possible.

### 4.3 Simulation Results

In this section, we discuss simple illustrative examples to learn certain performance aspects of a multihop network by exercising the model described in the previous sections. We consider a linear network of five nodes as shown in Figure 4.4, each node separated by a distance  $d$  from the other node. We label the nodes 1 through 5. These nodes form the vertices  $\mathcal{V}$  of a complete graph  $\mathcal{G}$ . Thus, there are  ${}^5C_2 = 10$  links in the network. Each node may be able to transmit to any other node in the network in just one hop. We assume that the transmissions are half duplex, i.e., nodes cannot transmit and receive simultaneously. Because of the half duplex constraint, the links that transmit in any slot should not contain any common vertex. The number of such sets of links is given by the number of edge independent sets of the graph  $\mathcal{G}$ . Thus, the number of transmission modes in this network is equal to the number of non-trivial matchings in the graph  $\mathcal{G}$ , i.e., 25.

We consider the physical layer to be a Gaussian interference channel. We run

the experiments assuming a bandwidth of 20 MHz. The wireless propagation model is taken to be the wideband PCS microcell model [68, Chapter 4] with a path loss exponent equal to 2.8. The path loss at a distance  $d$  meters away from the transmitter is given by  $PL = 38 + 28 \log_{10} d$  dB. For example, the interference gains between the transmitter of node 1 and the receiver of node 2 in Figure 4.4 is by  $G_{21} = 10^{-3.8/d^{2.8}}$ . Using a receiver noise figure of 3 dB, the receiver noise power is  $\sigma^2 = -104$  dBm. The transmit powers for all transmissions are fixed at 100 mW. The variation of rates due to interference between the links are captured by the matrix  $\mathbf{C}$ .

### 4.3.1 Maximizing sum of session rates in the network

We consider a single session, originating at node 1 and ending at node 5. Note that there are  $2^3 = 8$  different routes in the network for this session, since the nodes 2, 3, 4 can either be a part of a route or not. The objective is to maximize the session rate in the network. We thus use the objective function  $U(\mathbf{f}) = \mathbf{1}^T \mathbf{f}$ . For any inter-node distance  $d$  between the nodes, we can calculate the SINR for links in every possible mode and then construct the matrix  $\mathbf{C}$  for a fixed transmit power. By solving the LP in (4.9), we obtain the routes and the schedule for the modes required to obtain the optimal flows in these routes.

Figure 4.5 shows the variation of session rate with increasing inter-node distance  $d$ . We compare the optimal value with the single hop and the four-hop flows. For small values of  $d$ , the direct hop is the most optimal route. This also results in the highest sum rate since the mode with the single link (1, 5) can be used. When  $d$  increases, there is a four fold increase in the length of direct hop link. Hence the flow between the OD pair decreases rapidly due to the path loss. As  $d$  increases further, the single hop link is no more optimal and the flow takes more than one hop to reach the destination. For very large values of  $d$ , the path loss is dominant and it is optimal to route to the nearest neighbor. Table 4.1 shows the set of routes taken along with the time schedules for the active modes of transmissions for some sample values of  $d$ .

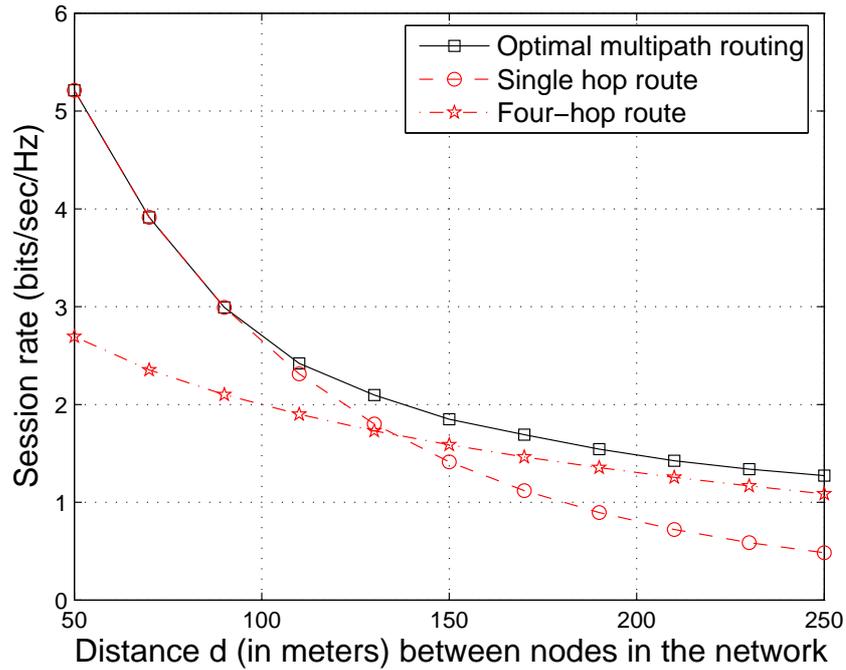
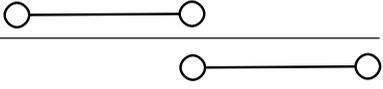
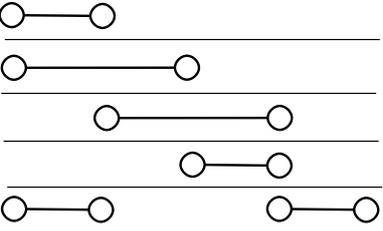
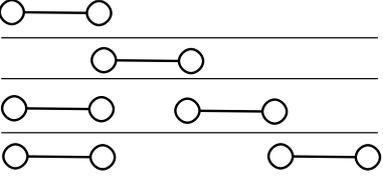


Figure 4.5: Variation of session rates as the inter-node distance  $d$  increases. Single hop routing is optimal for smaller values of  $d$  and the four-hop route is asymptotically optimal for large values of  $d$ .

#### 4.3.2 Performance with successive interference cancellation at the receiver

We now explore the benefit of successive interference cancellation (SIC) at the receiver node. We consider a session originating at node 1 and ending at node 5 as shown in Figure 4.4. We add additional complexity only at the receiver node for a session originating at node 1 and ending at node 5. In order to keep the scheduling complexity low, we make use of an ideal interference canceler to cancel one other interferer that transmits data to the receiver node 5. Thus, the receiver can decode two active flows simultaneously. The average rates in the links are determined by the order in which the links are decoded. The number of transmission modes in this case increases to 37. We term the link with a higher received SINR as the stronger link and the link with a lower received SINR as the weaker link. When the stronger link is decoded first, it suffers interference from the weaker link and all other active links and thus achieves a rate calculated using Equation (3.5). After the stronger link has been decoded, its

Table 4.1: Schedules for the transmission modes for different inter-node distances when single user decoding is used at the receiver

Inter-node separation distance $d$ (in meters)	Routes taken	Transmission modes	Mode activity time
10	$1 \rightarrow 5$		1.0
130	$1 \rightarrow 3 \rightarrow 5$		0.5, 0.5
170	$1 \rightarrow 2 \rightarrow 4 \rightarrow 5$ , $1 \rightarrow 3 \rightarrow 4 \rightarrow 5$		0.0251, 0.1668, 0.3622, 0.0910, 0.3549.
250	$1 \rightarrow 2 \rightarrow 3 \rightarrow 4 \rightarrow 5$		0.0151, 0.2929, 0.3649, 0.3272.

interference to the weaker link is completely canceled out at the receiver and so the weaker link suffers interference from all other active links except the stronger link. The rate achieved by the weaker link is then calculated using Equation (3.5).

Figure 4.6 shows the variation of session rates with increasing distance between the nodes of the network. When the stronger link is decoded first, followed by the weaker link, successive decoding provides gains that are worth pursuing. While the gains for low values of inter-node separation distance ( $0 < d \leq 50$ ) and higher inter-node separation distance ( $150 < d \leq 250$ ) are not significant, moderate values of inter-node separation distances ( $50 \leq d \leq 150$ ) yield gains that are more pronounced. Transmissions in this regime are more interference limited and hence interference cancellation helps. In the case when the weaker link is decoded first, and its interference is canceled to decode

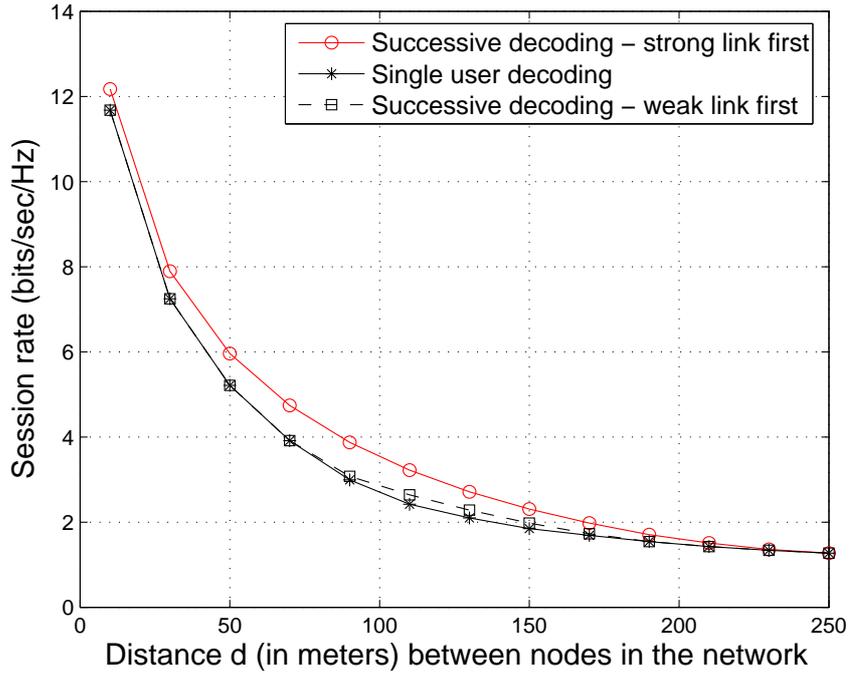


Figure 4.6: Comparison of throughput gains due to successive interference cancellation at the destination node

the stronger link, throughput gains are insignificant. This is because the average rates do not improve due to cancellation of weaker interferers.

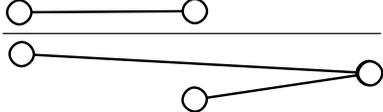
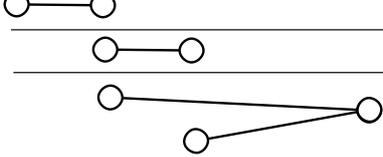
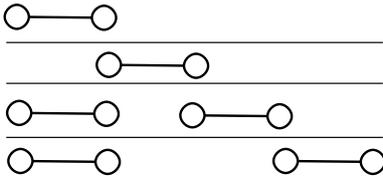
Table 4.2 shows the set of routes taken along with the time schedules for the active modes of transmissions for some sample values of  $d$ , when successive decoding is employed at the destination node.

### 4.3.3 Relation to Toumpis's work

The transmission modes used in our work are similar to the *transmission schemes* used in [64]. In [64], the author uses the tools of transmission schemes and rate matrices (described in detail below) to identify the capacity regions of wireless ad hoc networks under different transmission protocols. In the sequel, we identify the similarity of our work with [64] and show that a simple extension of [64] can achieve our results, but with more computational complexity.

First, we briefly describe the *rate matrices* used in [64]. For a network of  $n$  nodes,

Table 4.2: Schedules for the transmission modes for different inter-node distances when successive decoding (stronger link decoded first) is employed at the receiver

Inter-node separation distance $d$ (in meters)	Routes taken	Transmission modes	Fractions of time modes are active
10	$1 \rightarrow 2 \rightarrow 5$ , $1 \rightarrow 5$		0.0893, 0.9107.
70	$1 \rightarrow 3 \rightarrow 5$ , $1 \rightarrow 5$		0.305, 0.465.
230	$1 \rightarrow 2 \rightarrow 3 \rightarrow 5$ , $1 \rightarrow 2 \rightarrow 5$		0.2111, 0.1640, 0.5449.
250	$1 \rightarrow 2 \rightarrow 3 \rightarrow 4 \rightarrow 5$		0.0151, 0.2929, 0.3649, 0.3272.

the rate matrix  $R(\mathcal{S})$  of a transmission scheme  $\mathcal{S}$  is an  $n \times n$  square matrix with entries  $r_{ij}$  given by

$$r_{ij} = \begin{cases} r, & \text{if node } j \text{ receives information at rate } r \text{ with node } i \\ & \text{as the original information source} \\ -r, & \text{if node } j \text{ transmits information at rate } r \text{ with node } i \\ & \text{as the original information source} \\ 0, & \text{otherwise.} \end{cases} \quad (4.10)$$

The rate matrices capture all the information needed to describe the states of the system at a given time: namely, which nodes transmit or receive, at what rate, and from which nodes the data originate. A time division of transmission schemes is described by the weighted sum of the rate matrices with weights given by the time schedules. Given a

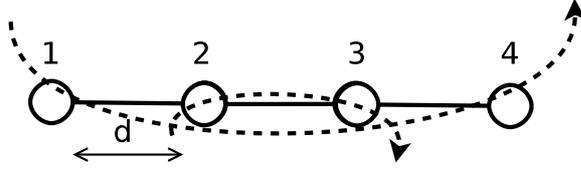


Figure 4.7: A four node linear network. We consider two flows sharing the available resources.

time-division schedule  $x_1, x_2, \dots, x_N$ , the rate matrix is given by  $R = \sum_{i=1}^N x_i R_i$ , where  $R_1, \dots, R_N$  are the rate matrices of the schemes  $\mathcal{S}_1, \dots, \mathcal{S}_N$ .

We illustrate the similarity of our work with [64] using an example. For ease of exposition, we consider a linear network of 4 nodes as shown in Figure 4.7. There are two flows, one originating from the node 1 and terminating at node 4 and the second originating at node 2 and terminating at node 3. In the simplest case, we would like to maximize the sum of the two flows in the network. We will show how the methodology in [64] can be used to perform the optimization and contrast with our framework.

We first enumerate the rate matrices as described by [64]. Let us denote the links  $(1, 2)$ ,  $(2, 3)$ ,  $(3, 4)$ ,  $(1, 3)$ ,  $(2, 4)$ ,  $(1, 4)$  as  $a$ ,  $b$ ,  $c$ ,  $d$ ,  $e$  and  $f$  respectively. We denote the physical layer rate in the link  $x$  in the absence of interference as  $r_u$  and in the presence of an interfering link  $v$  as  $r_{uv}$  where  $u, v \in \{a, b, c, d, e, f\}$ . Since, we use multihop routing with spatial reuse, we enumerate rate matrices for each end-to-end flow given the constraint that any node cannot simultaneously transmit and receive. The following is an enumeration of the rate matrices, assuming that a single flow originating from node 1 and terminating at node 4 exists in the network.

$$R_1 = \begin{pmatrix} -r_a & r_a & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}; R_2 = \begin{pmatrix} 0 & -r_b & r_b & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}; R_3 = \begin{pmatrix} 0 & 0 & -r_c & r_c \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix};$$

$$\begin{aligned}
R_4 &= \begin{pmatrix} -r_d & 0 & r_d & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}; R_5 = \begin{pmatrix} 0 & -r_e & 0 & r_e \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}; R_6 = \begin{pmatrix} -r_f & 0 & 0 & r_f \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}; \\
R_7 &= \begin{pmatrix} -r_{cd} & -r_{dc} & r_{cd} & r_{dc} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}; R_8 = \begin{pmatrix} -r_{ac} & r_{ac} & -r_{ca} & r_{ca} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}; \\
R_9 &= \begin{pmatrix} -r_{fb} & -r_{bf} & r_{bf} & r_{fb} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.
\end{aligned}$$

If there were only one flow originating at node 2 and terminating at node 3 in the network, the rate matrix is given by,

$$R'_1 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & -r_b & r_b & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

The above matrices were formed assuming that the only one flow exists in the network. If we allow both the flows in the network to coexist simultaneously, we need to include additional  $9 \times 1 = 9$  matrices to the number of matrices. Hence, we have  $9 + 1 + 9 = 19$  rate matrices in total.

In order to maximize the sum of flows shown in the Figure 4.7, we need to find the optimal time-division schedule  $\{x_1^*, x_2^*, \dots, x_N^*\}$  such that rate matrix  $R_{opt} = \sum_{i=1}^N x_i^* R_i$  (here  $N = 19$ ). We can formulate the optimization problem as a linear program as

shown below:

$$\max \quad r_{14} + r_{23} \quad (4.11a)$$

$$\text{subject to} \quad \sum_{i=1}^{19} x_i R_i \geq R_{opt}, \quad (4.11b)$$

$$\sum_{i=1}^{19} x_i = 1, \quad (4.11c)$$

$$x_i \geq 0, \quad i = 1, \dots, N, \quad (4.11d)$$

where,

$$R_{opt} = \begin{pmatrix} -r_{14} & 0 & 0 & r_{14} \\ 0 & -r_{23} & r_{23} & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

The above linear program in (4.11) can be easily solved by vectorizing the matrices. If we denote the vectorized version of the matrix  $R_i$  as  $\hat{\mathbf{r}}_i$  and the vectorized version of  $R_{opt}$  as  $\mathbf{r}_{opt}$ , we can rewrite (4.11) as

$$\max \quad r_{14} + r_{23} \quad (4.12a)$$

$$\text{subject to} \quad \sum_{i=1}^{19} x_i \hat{\mathbf{r}}_i \geq \hat{\mathbf{r}}_{opt}, \quad (4.12b)$$

$$\mathbf{1}^T \mathbf{x} = 1, \quad (4.12c)$$

$$\mathbf{x} \geq \mathbf{0}. \quad (4.12d)$$

If we collect the vectors  $\hat{\mathbf{r}}_i$ ,  $i = 1, \dots, 19$  in a  $16 \times 19$  matrix  $\hat{\mathbf{R}}$ , we have,

$$\max \quad r_{14} + r_{23} \quad (4.13a)$$

$$\text{subject to} \quad \hat{\mathbf{R}}\mathbf{x} \geq \hat{\mathbf{r}}_{opt}, \quad (4.13b)$$

$$\mathbf{1}^T \mathbf{x} = 1, \quad (4.13c)$$

$$\mathbf{x} \geq \mathbf{0}. \quad (4.13d)$$

The linear program (4.13) has  $v_1 = 2$  variables. The number of inequalities in

(4.13b) are 16, and (4.13d) contains 19 inequalities. The constraint set is of size  $c_1 = 16 + 1 + 19 = 36$ . The number of operations required to solve (4.13) is  $O(c_1^3)$  [69].

In comparison, if we use the framework described in Section 4.2 to solve the same problem, we can write the optimization program as below:

$$\max \quad f_1 + f_2 \quad (4.14a)$$

$$\text{subject to} \quad \mathbf{r} = \mathbf{C}\mathbf{x}, \quad (4.14b)$$

$$\mathbf{r} \geq \mathbf{A}(\mathbf{f}_1 + \mathbf{f}_2), \quad (4.14c)$$

$$\mathbf{1}^T \mathbf{x} = 1, \quad (4.14d)$$

$$\mathbf{x} \geq \mathbf{0}, \quad (4.14e)$$

$$\mathbf{f}_k \geq \mathbf{0}, \quad k = 1, 2. \quad (4.14f)$$

The LP (4.14) has  $v_2 = 2$  variables. The number of equalities and inequalities in (4.14b) and (4.14c) is 6. There are 9 inequalities in (4.14e) and 5 in (4.14f). Hence the size of the constraint set is  $c_2 = 6 + 6 + 1 + 9 + 5 = 27$ . The number of operations required to solve (4.14) is  $O(c_2^3)$ .

Both (4.13) and (4.14) are linear programs but (4.13) has a higher complexity since it has a bigger size of the constraint set. In contrast, the LP (4.14) using our framework has a lower complexity since we decouple the physical layer and the routing layers in (4.14b) and (4.14c). LPs using the framework in [64] grows exponentially faster for increasing number of end-to-end flows and could limit computations to very small networks.

#### 4.4 Conclusion

In this work, we presented a cross-layer optimization framework for scheduling rates and routing flows for end-to-end sessions in a wireless network. If the link gains are known, the spectrum server provides the schedule that are a collection of time shared transmission modes to achieve the end-to-end rates. The framework applies to a wide variety of physical layer schemes in which the rates of individual links depend on a single

transmitting configuration of links in the network. We illustrated the variation of the session rates and the routes with the distance between the nodes with and without successive decoding at the destination node. The resultant optimal routes and the schedule of transmission modes give us an idea of how we can select routes when the network operates in a decentralized way. Gains from successive decoding, when viewed from the end-to-end throughput perspective do not seem worthy enough. Finally, we observe that, even for a simple example setting of a linear network, the routing decisions can be complicated depending on the various rates in the links that can be obtained by different forms of signal processing in the physical layer.

## Chapter 5

### Conclusion

The results obtained in the first part of this dissertation gives some insights to practical benefits to relaying in interference limited wireless networks. Chapter 2 dealt with relaying in downlink cellular systems. We reported power savings and throughput improvement when half-duplex relays are deployed. Although the work provides insights into the order of gains due to relay deployments, further work is needed to assess the exact gains in a real cellular network because the simulations were conducted over idealized set-up. Some of the following issues require future work:

#### 1. Propagation models for cellular relay networks

Experimental studies for propagation models for cellular relay systems are currently being conducted. However, since the physical specifications of relays are not finalized, there are many empirical models in the literature. The propagation characteristics could be completely different if the relays are placed above roof-tops with line-of-sight communication to the base station, or if they are placed below roof-tops without line-of-sight communication to the base station. A standard model is yet to be specified.

#### 2. Practical coding schemes for implementing collaborative relaying

There is no known practical coding scheme that implements CPA scheme described in Chapter 2. There have been studies to show the practicality of such schemes [70]. Some attempts have been made [71–73], but there are quite a few open questions [74].

### 3. Low complexity schemes to find optimal order of base transmission

In the PC-CPA scheme, the base station employs a simple power control scheme to target the users in order to satisfy the desired common rate. There could arise a question as to why not target the relays? Or, depending upon the channel characteristics the set of users and relays could be partitioned into two groups. The base stations can target one group to get the desired target rate. The complexity of these schemes grow exponentially and there are very few special cases where these are addressed. A similar issue is the case when the transmission modes grow exponentially in Chapter 3.

In the second part of the dissertation, we studied centralized scheduling strategies. One of the important issues in that study is to find distributed algorithm to find the good transmission modes of operation. The column generation approach followed in [61], is a provably low-complex method to find good transmission modes. However, the column generation technique is still a centralized approach. For distributed approaches, we had explored a random access method in [54]. A recent study [75] proposes an adaptive CSMA scheduling that can distributively achieve the maximal throughput [76]. However, for an physical interference model introduced in [54], a distributed scheduling algorithm is still unknown. A future work is to find out ways to get good transmission modes and eventually find a distributed solution to the scheduling problem and evaluated how well it performs when compared to the centralized solution.

## References

- [1] Multi-hop relay system evaluation methodology. <http://ieee802.org/16>.
- [2] Andrea Goldsmith. *Fundamentals of Wireless Communication*. Cambridge University Press, Cambridge, UK, 2005.
- [3] A. J. Viterbi. *CDMA: Principles of Spread Spectrum Communication*. Addison Wesley, 1995.
- [4] Sergio Verdù. *Multuser Detection*. Cambridge University Press, Cambridge, UK, 1998.
- [5] R. H. Etkin, D. N. C. Tse, and Hua Wang. Gaussian interference channel capacity to within one bit. *IEEE Trans. Inf. Theory*, 54(12):5534–5562, Dec 2008.
- [6] N. Mandayam. Cognitive algorithms and architectures for open access to spectrum. *Conf. on the Economics, Technology and Policy of Unlicensed Spectrum*, May 2005. East Lansing, MI.
- [7] Y. Yang, H. Hu, J. Xu, and G. Mao. Relay technologies for WiMAX and LTE-advanced mobile systems. *IEEE Commun. Magazine*, 47(10):100–105, October 2009.
- [8] L. Le and E. Hossain. Multihop cellular networks: Potential gains, research challenges, and a resource allocation framework. *IEEE Commun. Magazine*, 45(9):66–73, September 2007.
- [9] H. Viswanathan and S. Mukherjee. Performance of cellular networks with relays and centralized scheduling. *IEEE Trans. Wireless Commun.*, 4(5):2318–2328, Sept 2005.
- [10] O. Oyman, J. N. Laneman, and S. Sandhu. Multihop relaying for broadband wireless mesh networks: from theory to practice. *IEEE Commun. Magazine*, 45(11):116–122, Nov 2007.
- [11] Ozgur Oyman. Opportunistic scheduling and spectrum reuse in relay-based cellular OFDMA networks. In *Proc. Globecom*, 2007. Washington DC.
- [12] Charles E. Perkins. *Ad Hoc Networking*. Addison-Wesley, 2001.
- [13] E. C. van der Meulen. *Transmission of Information in a T-terminal discrete memoryless channel*. PhD thesis, Univ. of California, Berkeley, CA, 1968.
- [14] T. M. Cover and A. A. El Gamal. Capacity theorems for the relay channel. *IEEE Trans. Inf. Theory*, 25(5):572–584, Sept 1979.

- [15] G. Kramer, M. Gastpar, and P. Gupta. Cooperative strategies and capacity theorems for relay networks. *IEEE Trans. Inf. Theory*, 51:3037–3063, Sept 2005.
- [16] M. A. Khojastepour, A. Sabharwal, and B. Aazhang. On the capacity of ‘cheap’ relay networks. In *Proc. 37th CISS*, 2003. Baltimore, MD.
- [17] G. Kramer. Models and theory for relay channels with receive constraints. In *Proc. of Allerton Conf. on Comm., Control and Computing*, 2004. UIUC, IL.
- [18] D. Chen, M. Haenggi, and J. N. Laneman. Distributed spectrum-efficient routing algorithms in wireless networks. *IEEE Trans. Wireless Commun.*, 7(12):5297–5305, Dec 2008.
- [19] M. Sikora, J. N. Laneman, M. Haenggi, Jr. D. J. Costello, and T. E. Fuja. Bandwidth- and power-efficient routing in linear wireless networks. *IEEE Trans. Inf. Theory*, 52(6):2624–2633, June 2006.
- [20] P. Bender, P. Black, M. Grob, R. Padovani, N. Sindhushayana, and A. Viterbi. CDMA/HDR: A bandwidth-efficient high speed wireless data service for nomadic users. *IEEE Commun. Magazine*, 38(7):70–77, July 2000.
- [21] D. Bertsekas and R. Gallager. *Data Networks*. Prentice-Hall, 1992.
- [22] Harmonized Contribution on 802.16j (Mobile Multihop Relay) Usage Models. <http://ieee802.org/16>.
- [23] R. Pabst, B. H. Walke, D. C. Schultz, P. Herhold, H. Yanikomeroglu, S. Mukherjee, H. Viswanathan, M. Lott, W. Zirwas, M. Dohler, H. Aghvami, D. D. Falconer, and G. P. Fettweis. Relay-based deployment concepts for wireless broadband radio. *IEEE Commun. Magazine*, 42(9):80–89, Sept 2004.
- [24] G. Foschini, A. Tulino, and R. Valenzuela. Performance comparison for basic relay systems. *Tech. Memo, Bell Labs NJ*, 2008.
- [25] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. New York:Wiley, 2006.
- [26] M. Katz and S. Shamai. Transmitting to colocated users in wireless ad hoc and sensor networks. *IEEE Trans. Inf. Theory*, 51:3540–3563, Oct 2005.
- [27] D. Blackwell, L. Breiman, and A. J. Thomasian. Capacity of a class of channels. *Annals of Mathematical Statistics*, 30:1229–1241, 1968.
- [28] I. Csiszar and P. Narayan. Capacity of the gaussian arbitrarily varying channel. *IEEE Trans. Inf. Theory*, 37:18–26, Jan 1991.
- [29] David Tse and Pramod Viswanath. *Fundamentals of Wireless Communication*. Cambridge University Press, Cambridge, UK, 2005.
- [30] K. Azarian, H. El Gamal, and P. Schniter. On the achievable diversity-vs-multiplexing tradeoff in half-duplex cooperative channels. *IEEE Trans. Inf. Theory*, 51:4152–4172, Dec 2005.

- [31] M. Chiang, P. Hande, T. Lan, and C. W. Tan. Power control in wireless cellular networks. *Foundations and Trends in Networking*, 4(2):381–533, July 2008.
- [32] N. Jacobsen. Practical cooperative coding for half-duplex relay channels. In *Proc. CISS 2009*, March 2009.
- [33] J. Mitola. *Cognitive radio: An integrated agent architecture for software defined radio*. PhD thesis, KTH Royal Institute of Technology, 2000.
- [34] Wei Zhang and K. Ben Letaief. Cooperative communications for cognitive radio networks. *Proceedings of the IEEE*, 97(5):878–893, May 2009.
- [35] M. Buddhikot, P. Kolodzy, S. Miller, K. Ryan, and J. Evans. DIMSUMNet: New directions in wireless networking using coordinated dynamic spectrum access. In *IEEE WoWMoM*, June 2005.
- [36] O. Ileri, D. Samardzija, T. Sizer, and N. Mandayam. Demand responsive pricing and competitive spectrum allocation via a spectrum policy server. In *Proc. IEEE DySPAN*, Nov 2005. Baltimore, MD.
- [37] P. Gupta and P. R. Kumar. Capacity of wireless networks. *IEEE Trans. Inf. Theory*, 46(2):388–404, Mar 2000.
- [38] K. Jain, J. Padhye, V. N. Padmanabhan, and L. Qiu. Impact of interference on multi-hop wireless network performance. In *ACM Mobicom*, Sept 2003. San Diego, CA.
- [39] B. Hajek and G. Sasaki. Link scheduling in polynomial time. *IEEE Trans. Inf. Theory*, 34(5):910–917, Sept 1988.
- [40] G. Middleton, B. Aazhang, and J. Lilleberg. A flexible framework for polynomial-time resource allocation in multiflow wireless networks. In *Proc. of Allerton Conf. on Comm., Control and Computing*, 2009. UIUC, IL.
- [41] T. Elbatt and A. Ephremides. Joint scheduling and power control for wireless ad-hoc networks. *IEEE Trans. Wireless Commun.*, 3(1):74–85, Jan 2004.
- [42] G. Foschini and Z. Miljanic. A simple distributed autonomous power control algorithm and its convergence. *IEEE Trans. Veh. Technol.*, 42(4):641–646, Nov 1993.
- [43] R. L. Cruz and A. V. Santhanam. Optimal routing, link scheduling and power control in multi-hop wireless networks. In *Proc. IEEE Infocom*, 2003.
- [44] R. Bhatia and M. Kodialam. On power efficient communication over multi-hop wireless networks: Joint routing, scheduling and power control. In *Proc. IEEE Infocom*, 2004.
- [45] L. Chen, S. Low, M. Chiang, and J. Doyle. Optimal cross-layer congestion control, routing and scheduling design in ad hoc wireless networks. In *Proc. IEEE INFOCOM*, 2006.
- [46] X. Lin, N. B. Shroff, and R. Srikant. A tutorial on cross-layer optimization in wireless networks. *IEEE J. Sel. Areas Commun.*, 24(8):1452–1463, Aug 2006.

- [47] R. L. Cruz and A. V. Santhanam. Optimal link scheduling and power control in CDMA multihop wireless networks. In *IEEE Globecom*, pages 52–56, 2002.
- [48] M. Chiang, S. H. Low, A. R. Calderbank, and J. C. Doyle. Layering as optimization decomposition: A mathematical theory of network architectures. *Proceedings of the IEEE*, 95(1):255–312, Jan 2007.
- [49] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.
- [50] B. Radunovic and J. Y. L. Boudec. Rate performance objectives of multihop wireless networks. *IEEE Trans. Mobile Comput.*, 3(4):334–349, Oct.-Dec 2004.
- [51] F. Kelly. Charging and rate control for elastic traffic. *Euro. Trans. Telecommun.*, 8:33–37, Jan/Feb 1997.
- [52] D. Bertsekas and J. Tsitsiklis. *Parallel and Distributed Computation: Numerical Methods*. Athena Scientific, 1997.
- [53] A. Ephremides. Energy concerns in wireless networks. *IEEE Wireless Commun.*, 9(4):48–59, Aug 2002.
- [54] J. Singh, C. Raman, R. Yates, and N. Mandayam. Random access for variable rate links. In *Proc. MILCOM 2006*, Oct 2006. Washington D.C.
- [55] E. Arıkan. Some complexity results about packet radio networks. *IEEE Trans. Inf. Theory*, 30(4):681–685, July 1984.
- [56] P. Varbrand, D. Yuan, and P. Bjorklund. Resource optimization of spatial TDMA in ad hoc radio networks: A column generation approach. In *Proc. IEEE Infocom*, 2003.
- [57] X. Lin and N. B. Shroff. Joint rate control and scheduling in multihop networks. In *43rd IEEE Conference on Decision and Control*, December 2004. Paradise Island, Bahamas.
- [58] S. Sarkar, P. Chaporkar, and K. Kar. Fairness and throughput guarantees with maximal scheduling in multi-hop wireless networks. In *Proc. Workshop on Modelling and Optimization in Mobile, Ad Hoc and Wireless Networks*, Apr 2006. Boston, MA.
- [59] S. Sarkar and L. Tassiulas. End-to-end bandwidth guarantees through fair local spectrum share in wireless ad hoc networks. *IEEE Trans. Autom. Control*, 50(9):1246–1259, Sept 2005.
- [60] M. Kodialam and T. Nandagopal. Characterizing achievable rates in multi-hop wireless networks: the joint routing and scheduling problem. In *ACM Mobicom*, Sept 2003. San Diego, CA.
- [61] M. Johansson and L. Xiao. Cross-layer optimization of wireless networks using nonlinear column generation. *IEEE Trans. Wireless Commun.*, 5(2):435–445, Feb 2006.

- [62] C. Raman, R. Yates, and N. Mandayam. Scheduling variable rate links via a spectrum server. In *Proc. IEEE DySPAN*, 2005. Baltimore, MD.
- [63] R. Yates, C. Raman, and N. Mandayam. Fair and efficient scheduling of variable rate links via a spectrum server. In *Proc. IEEE ICC*, Jun 2006. Istanbul, Turkey.
- [64] S. Toumpis and A. Goldsmith. Capacity regions of wireless ad hoc networks. *IEEE Trans. Wireless Commun.*, 2(4):736–748, July 2003.
- [65] S. Mukherjee and H. Viswanathan. Throughput-range tradeoff of wireless mesh backhaul networks. *IEEE J. Sel. Areas Commun.*, 24(3):593–602, Mar 2006.
- [66] D. Bertsimas and J. Tsitsiklis. *Introduction to Linear Optimization*. Athena Scientific, 1997.
- [67] J. Mo and J. Walrand. Fair end-to-end window-based congestion control. *IEEE/ACM Trans. Netw.*, 8(5):556–567, Oct 1997.
- [68] T. S. Rappaport. *Wireless Communications: Principles and practice*. Prentice Hall, 2001.
- [69] C. C. Gonzaga. On the complexity of linear programming. *Resenhas - IME-USP Journal, special issue dedicated to Paul Erdos*, 2(2):197–207, 1995.
- [70] G. J. Foschini. Some basic relay arrangements with an application to cellular networks. *Workshop in Honor of David Goodman*, October 2009. Brooklyn, NY.
- [71] G. Kramer. Distributed and layered codes for relaying. In *Proc. 39th Asilomar Conference on Signals, Systems and Computers*, November 2005.
- [72] A. Chakrabarti, A. de Baynast, A. Sabharwal, and B. Aazhang. LDPC parity check codes for the relay channel. *IEEE J. Sel. Areas Commun.*, 25(2):280–291, Feb 2007.
- [73] A. Chakrabarti, E. Erkip, A. Sabharwal, and B. Aazhang. Code designs for cooperative communication. *IEEE Signal Processing Magazine*, 24(5):16–26, Sept. 2007.
- [74] G. Kramer. Cooperative communications and coding. In *Turbo coding symposium*, 2008. EPFL, Lauzanne.
- [75] L. Jiang and J. Walrand. A distributed csma algorithm for throughput and utility maximization in wireless networks. In *Proc. of Allerton Conf. on Comm., Control and Computing*, 2008.
- [76] C. Joo, X. Lin, and N. Shroff. Understanding the capacity region of the greedy maximal scheduling algorithm in multihop wireless networks. *IEEE/ACM Trans. Netw.*, 17(4):1132–1145, Aug. 2009.

## Curriculum Vitae

### Chandrasekharan Raman

#### EDUCATION

- 2010** Ph.D. (Electrical & Computer Engineering), Rutgers – The State University of New Jersey, US
- 2008** M.S. (Electrical & Computer Engineering), Rutgers – The State University of New Jersey, US
- 2003** M.E. (Telecommunications Engineering), Indian Institute of Science, Bangalore, India
- 2001** B.E. (Electronics & Communications Engineering), College of Engineering, Guindy, Anna University, Chennai, India

#### EMPLOYMENT/CAREER

- 2005–2010** Graduate Research Assistant, WINLAB, Rutgers - The State University of New Jersey
- 2007–2008** Visiting Student Researcher, Wireless Communication Research Labs, Bell Labs, Alcatel-Lucent, Holmdel NJ
- 2006** Interim Engineering Intern, Corporate R & D, Qualcomm, San Diego CA
- 2004–2005** Graduate Teaching Assistant, Electrical & Computer Engineering Dept., Rutgers - The State University of New Jersey
- 2003–2004** R & D Engineer, Tejas Networks India Limited, Bangalore, India

#### PUBLICATIONS

- 2010** C. Raman, G. J. Foschini, R. A. Valenzuela, R. D. Yates and N. B. Mandayam, “Collaborative Relaying in Downlink Cellular Systems,” book chapter in *Cooperative Cellular Wireless Networks*, V. Bhargava, E. Hossain and D. I. Kim (Eds.), Cambridge University Press 2010.
- 2009** C. Raman, G. J. Foschini, R. A. Valenzuela, R. D. Yates and N. B. Mandayam, “Power Savings from Half-Duplex Relaying in Downlink Cellular Systems,” in *Proc. Allerton Conference 2009*, Monticello, IL.

- 2007** C. Raman, J. Kalyanam, I. Seskar, and N. B. Mandayam, "Distributed Spatio-Temporal Spectrum Sensing: An Experimental Study," in *Proc. Asilomar Conf. SSC*, Pacific Grove, CA.
- 2007** C. Raman, J. Singh, R. D. Yates and N. B. Mandayam, "Scheduling in Cognitive Networks: Scheduling variable rate links - centralized and decentralized approaches," book chapter in *Cognitive Wireless Networks*, Frank H.P. Fitzek, Marcos D. Katz (Eds.) 2007, Springer, Signals and Communication series.
- 2006** J. Singh, C. Raman, R. D. Yates and N. B. Mandayam, "Random Access for Variable Rate Links," in *Proc. MILCOM 2006*, Washington D.C.
- 2006** R. D. Yates, C. Raman and N. B. Mandayam, "Fair and Efficient Scheduling of Variable Rate Links via a Spectrum Server," in *Proc. International Communications Conference 2006*, Istanbul, Turkey.
- 2006** C. Raman, R. D. Yates and N. B. Mandayam, "Cross-layer Scheduling of End-to-End Flows using a Spectrum Server" in *Proc. CISS 2006*, Princeton, NJ.
- 2005** C. Raman, R. D. Yates and N. B. Mandayam, "Scheduling of Variable Rate Links via a Spectrum Server," in *Proc. DySpan 2005*, Baltimore, MD.