Exploiting the Benefits of Cooperation in Wireless Networks

D I S S E R T A T I O N

for the Degree of

Doctor of Philosophy (Electrical Engineering)

Sha Hua

Jan 2013
Exploiting the Benefits of Cooperation in Wireless Networks

DISSERTATION

Submitted in Partial Fulfillment
of the REQUIREMENTS for the

Degree of

DOCTOR OF PHILOSOPHY (Electrical Engineering)

at the

POLYTECHNIC INSTITUTE OF NEW YORK UNIVERSITY

by

Sha Hua

Jan 2013

Approved:

________________________
Department Head

________________________
Date

Copy No. ___
Approved by the Guidance Committee:

Major: Electrical and Computer Engineering

Shivendra S. Panwar
Professor of
Electrical and Computer Engineering

Yong Liu
Associate Professor of
Electrical and Computer Engineering

Sundeep Rangan
Associate Professor of
Electrical and Computer Engineering

Minor: Computer Science

Justin Cappos
Assistant Professor of
Computer Science and Engineering
Microfilm or other copies of this dissertation are obtainable from

UMI Dissertations Publishing
Bell & Howell Information and Learning
300 North Zeeb Road
P.O.Box 1346
Ann Arbor, Michigan 48106-1346
VITA

**Sha Hua** was born in China on June 7, 1985. He received the B.E. degree in Electronics and Information Engineering from Huazhong University and Science and Technology in 2007. Since August 2007, he has been a Ph.D. candidate at Electrical and Computer Engineering Department in Polytechnic Institute of New York University, Brooklyn, NY 11201, under the supervision of Prof. Shivendar S. Panwar. During the period of May 2008 to September 2008, and May 2010 to September 2010, he interned at the Corporate Research, Thomson Inc. (now Technicolor Inc.). He was also a fixed access research intern at Bell Labs, Alcatel-Lucent, from June 2011 to August 2011. During his thesis study, he published five technical papers and had one patent pending. He has conducted research in the fields of wireless cooperative networking, cognitive radio and video delivery over wireless networks.
To my wife Xuejun and our parents,
for their love and support.
ACKNOWLEDGEMENT

First, I would like to gratefully and sincerely thank my thesis advisor Professor Shivendra S. Panwar for his invaluable support, inspiration, patience and guidance throughout my research at Polytechnic Institute of NYU. I have learned a lot from his insightful comments and constructive criticism, which corrected my faltering steps, and will continue to benefit my future career. I am also indebted to other committee members of this dissertation. Professor Yong Liu, Professor Sundeep Rangan and Professor Justin Cappos have constantly given me research advice. Their way of tackling problems and thinking in a higher plane deeply influenced me.

I would also like to thank Dr. Hang Liu, Dr. Yang Guo and Dr. Pei Liu, who worked with me closely on many projects through my research. The discussions with them are always inspiring and helpful. I am also very grateful to my colleagues at Technicolor Corporate Research and Bell Labs, Alcatel-Lucent: Dr. Mingquan Wu, Dr. Yunfei Zheng, Edward Grinshpun for their warm-hearted help and insightful technical discussion. Thanks to the members of the Network Management Lab at Polytechnic: Prof. Yanming Shen, Dr. Shunyuan Ye, Dr. Chun Nie, Shu Luo, Ayaskant Rath, Zizhong Cao, Tingting Lu, Georgios Kyriakou, Sanjay Goyal, Seyed Amir Hossein Hosseini, Nicolas Barati and Fraida Fund, for their collaboration, discussion and activities, which brightened my graduate life. I also want to thank other members and friends at Polytechnic Institute of NYU: Dr. Chao Liang, Dr. Chun Nie, Dr. Zhengye Liu, Dr. Zhan Ma, Dr. Hao Hu, Dr. Jinru He, Dr. Chao Zhang, Xiwang Yang, Yang Xu, Guibin Tian, Siming Wang. Because of all of them, my life here is much more enjoyable.

In particular, I want to thank my wife, Dr. Xuejun Zhuo, for her continuous caring and love. The days we were both working towards the doctoral degree are precious memory to me. Without her, the achievements of my thesis are impossible. Finally, I would like to thank our parents for their persistent support and encouragement.

This work is supported by National Science Foundation, the New York State Center for Advanced Technology in Telecommunications (CATT) and the NSF Wireless Internet
Center for Advanced Technology (WICAT) at Polytechnic Institute of New York University, Brooklyn, NY, USA.
AN ABSTRACT

Exploiting the Benefits of Cooperation in Wireless Networks

by

Sha Hua

Advisor: Shivendra S. Panwar

Submitted in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy (Electrical Engineering)

Jan 2013

Given scarce spectrum resources and the ever-increasing data demands of the mobile users, methods to improve the spectrum efficiency is an active research area. Cooperation is one of the promising solutions that can lead to a significant performance improvement of the wireless system. The main objective of this thesis is to leverage the advanced cooperation techniques in typical wireless systems.

Cellular networks are viewed as the most successful deployment of wireless networks to provide ubiquitous voice and data communications to the mobile users. Video broadcast/multicast service over cellular networks is a popular application, but supporting this is a challenging task, which requires high-bandwidth and low latency. We propose a novel scheme using scalable video coding (SVC) and user cooperation to improve the system-wide user video quality of experience. The users receiving more layers of video can forward the extra layers to the users with fewer layers. Practical protocols for discovering the helpers and efficiently distributing video content are designed. The cellular base station conducts a centralized resource allocation to maximize the overall video quality among all users in the multicast session.
Cognitive radio networks have recently been proposed as a new solution to the spectrum scarcity problem. It allows more efficient utilization of the licensed spectrum by allowing the unlicensed secondary users (SUs), equipped with frequency-agile cognitive radios, to dynamically access the licensed spectrum held by the primary users (PUs). We propose to utilize multiple antennas on the SUs to cooperatively relay the traffic for the PUs while concurrently accessing the same channel to transmit their own traffic. Our scheme, named MIMO-CCRN, is designed by considering both the temporal and spatial domains to improve spectrum efficiency. We also consider two practical network scenarios. One scenario has a single primary link and multiple secondary relays, which we analyze using a Stackelberg game model. Another scenario, with multiple PUs and SUs, is analyzed by finding an optimal matching. In all scenarios, MIMO-CCRN is verified to achieve higher utility for both PUs and SUs.

Cross-layer designs are used in the above work to identify the benefits of cooperation. Either physical layer cooperation based on multiple antennas or network layer cooperation by multi-hop forwarding improves system performance. Additionally, we investigate cooperation among different wireless service providers. Specifically, we study the benefits of sharing infrastructure and spectrum among multiple cellular operators. A multi-cell analytical model using stochastic geometry is provided to analyze the average user throughput under cellular operator cooperation. With both analytical and simulation results, we show large capacity gains exist, providing a strong incentive for operators to cooperate.
List of Publications


Contents

| List of Figures | xiii |
| List of Tables | xv |

1 Introduction

1.1 Motivation for This Dissertation
   1.1.1 Multi-hop Cooperative Forwarding .......................... 1
   1.1.2 PHY Layer Cooperative Communications .................. 4
   1.1.3 Cooperation among Network Service Providers ............ 6

1.2 Dissertation Outline ........................................... 8

2 Scalable Video Multicast in Hybrid Cellular/Ad-Hoc Networks 10

2.1 Background and Motivation ...................................... 10
2.2 Scalable Video Broadcast/Multicast Service (SV-BCMCS) .... 14
   2.2.1 System Architecture .................................. 14
   2.2.2 Optimal Resource Allocation in Layered Video Multicast ... 16
   2.2.3 Dynamic Programming Algorithm Solution ................. 19

2.3 SV-BCMCS: Protocol Design .................................. 21
   2.3.1 Greedy Helper Discovery Protocol ...................... 21
   2.3.2 SV-BCMCS Relay Routing Protocol ..................... 23

2.4 Analysis of the Gain of Ad-hoc Relay ......................... 25
   2.4.1 Distance Gain Using One-Hop Relay .................... 26
   2.4.2 Distance Gain Using Two-Hop Relay .................... 28
   2.4.3 Impact of ad-hoc wireless relay to user performance ... 30
   2.4.4 Numerical Results Using the Analytical Model .......... 31

2.5 Performance Evaluation ......................................... 33
   2.5.1 Simulation Settings .................................... 33
   2.5.2 Stationary Scenarios .................................. 35
   2.5.3 Mobile Scenarios ...................................... 38
   2.5.4 Tradeoffs between the Base Layer Rate and the Overall Performance .... 40

2.6 Summary .......................................................... 41
## 3 Exploiting Multiple Antennas in Cooperative Cognitive Radio Networks

3.1 Background and Motivation .................................................. 44
3.2 Preliminaries: MIMO Characteristics .................................... 49
  3.2.1 Zero-Forcing Beam-forming ........................................... 49
  3.2.2 ZFBF in multi-user MIMO scenarios ............................... 50
  3.2.3 Remarks on employing ZFBF ........................................... 52
3.3 The Case of One Primary Link and Multiple SUs ......................... 52
  3.3.1 System Model ............................................................. 52
  3.3.2 Link Data Rate Analysis .............................................. 56
  3.3.3 Problem Formulation .................................................. 57
  3.3.4 Beyond Two Antennas ................................................ 58
  3.3.5 Game Theory Analysis ............................................... 59
3.4 The Case of MIMO-empowerd Femtocells .................................. 63
  3.4.1 System Model ............................................................. 64
  3.4.2 Coalition Formation ................................................... 67
3.5 Performance Evaluation ..................................................... 70
  3.5.1 Simulations for the Case of One Primary Link ...................... 70
  3.5.2 Simulations for the MIMO Femtocell Case .......................... 74
3.6 Conclusion ......................................................................... 76

## 4 Cooperation among Wireless Service Providers

4.1 Background and Motivation .................................................. 78
4.2 Analytical Modeling ............................................................ 81
  4.2.1 Average User Ergodic Rate ............................................ 82
  4.2.2 Average User Throughput ............................................. 84
  4.2.3 Special Case: Cooperation among Same-Size Operators ......... 85
4.3 An OFDMA-based Multi-Cell System ...................................... 86
4.4 Performance Evaluation ..................................................... 89
  4.4.1 Numerical Evaluations ................................................ 90
  4.4.2 Performance with Real BS Locations ............................... 93
4.5 Conclusion ......................................................................... 96

## 5 Conclusion and Future Work

5.1 Summary of Major Contributions ......................................... 97
5.2 Future Work ...................................................................... 99
  5.2.1 Content Delivery over Hybrid Cellular/Ad-hoc Networks ........ 99
  5.2.2 Cooperative Cognitive Radio Networks ............................ 100
  5.2.3 Cooperation among Wireless Service Providers ................. 100

**Bibliography** ................................................................. 102
List of Figures

1.1 A typical ad-hoc network with multi-hop transmission from source to destination. ................................................. 2
1.2 An example of cooperative communications ................................................................. 5
1.3 An example of cooperation among two different co-located network providers. The nodes of one operator can forward the traffic for the nodes in the other operator, thus resulting in a performance improvement of both providers. ............................................. 7
2.1 Coding architecture of of temporal scalable video in MPEG-4 SVC and H.264/AVC. ................................................................. 13
2.2 Architecture of SV-BCMCS over a hybrid network (assuming three layers of video content). ............................................. 15
2.3 Greedy Helper Discovery Protocol and Relay Routing Protocol. ........................................... 23
2.4 The illustrative example for the effective distance of the users. ........................................... 25
2.5 The one-hop (left) and two-hop (right) ad-hoc relay analysis for an arbitrary user. ...................... 27
2.6 Reference figure for positions of the user, helper 1 and helper 2. We label the circles as $C_A$, $C_B$, $C_C$ and $C_D$, with $C_A$ and $C_B$ centered at the user and helper 1, $C_C$ and $C_D$ centered at the Base Station. I, II and III indicate the three different patterns that $C_D$ intersect with $C_A$ and $C_B$. ........................................... 29
2.7 The one-hop and two-hop Distance Gain for nodes with different distances to the base station. ........................................... 32
2.8 Impact of number of users (a) Video sequence Mobile (b) Video sequence Football. ............................................... 36
2.9 Impact of number of users (Video sequence Bus). ............................................................... 37
2.10 Impact of TTL on the average PSNR. ............................................................................. 38
2.11 Impact of maxspeed on the average PSNR. ................................................................. 39
2.12 Impact of base layer encoding rate. ................................................................................. 39
2.13 Tradeoffs between the base layer rate and the overall performance in SV-BCMCS. ................................................................. 42
3.1 An illustration of the conventional CCRN scheme. ................................................................. 46
3.2 The motivating scenario for MIMO-CCRN. ........................................................................ 47
3.3 Transmission of two streams: a $2 \times 2$ MIMO channel. ..................................................... 49
3.4 Transmission of two streams: a multi-user MIMO scenario. ............................................... 51
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.5</td>
<td>System architecture and frame structure of MIMO-CCRN in the case of one primary link and multiple secondary user pairs.</td>
<td>54</td>
</tr>
<tr>
<td>3.6</td>
<td>A typical scenario for MIMO-CCRN in cellular networks with multiple MIMO-empowered femtocells.</td>
<td>64</td>
</tr>
<tr>
<td>3.7</td>
<td>Spectrum leasing between a primary link and a secondary network led by a Femto-SAP.</td>
<td>65</td>
</tr>
<tr>
<td>3.8</td>
<td>Primary link’s utility for different schemes.</td>
<td>71</td>
</tr>
<tr>
<td>3.9</td>
<td>Impact of the number of relays.</td>
<td>72</td>
</tr>
<tr>
<td>3.10</td>
<td>The utilities achieved by PUs and SUs.</td>
<td>74</td>
</tr>
<tr>
<td>3.11</td>
<td>Average utility of the secondary pairs.</td>
<td>75</td>
</tr>
<tr>
<td>3.12</td>
<td>Average utility of the UEs vs. the number of SAPs</td>
<td>76</td>
</tr>
<tr>
<td>3.13</td>
<td>Average utility of the secondary networks vs. the number of SAPs</td>
<td>76</td>
</tr>
<tr>
<td>4.1</td>
<td>Motivating scenario of cooperation between cellular operators.</td>
<td>79</td>
</tr>
<tr>
<td>4.2</td>
<td>A typical OFDMA frame in a temporal-frequency domain.</td>
<td>87</td>
</tr>
<tr>
<td>4.3</td>
<td>A 20 × 20 km view of the current BS deployment by two major cellular operators in the Washington D.C. area.</td>
<td>90</td>
</tr>
<tr>
<td>4.4</td>
<td>The impact of BS density.</td>
<td>91</td>
</tr>
<tr>
<td>4.5</td>
<td>The impact of user density.</td>
<td>91</td>
</tr>
<tr>
<td>4.6</td>
<td>The impact of bandwidth.</td>
<td>92</td>
</tr>
<tr>
<td>4.7</td>
<td>Network performance under OFDMA-based system.</td>
<td>94</td>
</tr>
<tr>
<td>4.8</td>
<td>User’s average throughput CDF.</td>
<td>95</td>
</tr>
</tbody>
</table>
List of Tables

2.1 Notations ................................................. 16
2.2 Rate (kb/s) and PSNR (dB) values of all the layers for the three SVC encoded video sequences ........................................ 34
2.3 Impact of Reconfiguration Interval on the Average PSNR ($maxspeed=10$ m/s) ........................................ 40
3.1 Optimal $\alpha^*$ versus the distance $d.$ ........................................ 72
4.1 OFDMA System Parameters ........................................ 94
Chapter 1

Introduction

1.1 Motivation for This Dissertation

The rapid growth in demand for high-speed and ubiquitous mobile communication requires a drastic improvement in the coverage, link capacity and latency of wireless access networks. A recent study from Cisco predicts that mobile data traffic will skyrocket by a factor between 25X to 50X by 2015 [1]. Cooperative networking is considered as a promising technique to improve the throughput and extend the wireless coverage. It is not only built upon physical (PHY) layer and medium access control (MAC) layer, but also spans the network layer and influences the performance of applications running above it. Therefore, cross-layer network design approaches are needed to realize high-performance systems. Moreover, different wireless service providers, holding orthogonal resources, can also collaborate with each other by jointly operating their respective networks. Such cooperation in network planning benefits the end users.

1.1.1 Multi-hop Cooperative Forwarding

As the wireless signal strength decreases drastically with the increase of the distance between the transmitter and the receiver, in order to enable end-to-end communications between the users far away from each other, multi-hop transmission is a necessity. This architecture is similar to the public Internet deployment that consists of thousands of routers that
are inter-connected and forward the traffic from the source to the destination. In the wireless domain, large-scale cellular networks and satellite networks are deployed to achieve the goal of long-distance communications [2]. However, a huge amount of expenditure is required to construct and maintain the infrastructure, including base stations, core networks and satellites. By contrast, when the infrastructure is not present, multi-hop ad-hoc networks can be used for data transmissions and at much lower cost.

An ad-hoc wireless network is a collection of wireless mobile nodes that self-configure to form a network without the aid of any established infrastructure [3], as shown in Figure 1.1. In ad-hoc networks, each node may communicate directly to the nodes in its communication range. When two nodes are not directly connected, the intermediate nodes will forward their traffic, acting as the routers. Therefore, ad-hoc network is intrinsically based on the nodes’ cooperation. Due to its advantages of flexibility, rapid deployment and re-configuration and robustness, it was originally invented for military applications. Later on, it evolves into various forms of networks, including wireless sensor networks (WSN) [4], wireless mesh networks [5], and Vehicular Ad hoc Networks (VANET) [6]. Recently, there also appeared a new type of ad-hoc network, Delay Tolerant Networks (DTN), without instantaneous end-to-end connectivity [7]. What they have in common is that cooperation and multi-hop forwarding are the core functions of these networks.
Despite their simple form, the optimal design, performance and fundamental capabilities of the multi-hop network remain poorly understood, and has fueled intensive research activity in recent years. Researchers have attempted to study it from a protocol layering point of view. For example, the MAC layer design and analysis are discussed in [8–10], routing protocols are studied in [11, 12] and the performance of TCP is studied in [13, 14]. Besides, as many protocol design issues are intertwined, cross-layer approaches are also presented, which can be found in [15–17]. With these innovations, the performance of multi-hop network has improved. However, scalability and latency are still the fundamental issues. In a canonical work [18], it was proved that the throughput obtained by each user in a multi-hop network for a randomly chosen destination is \( \Theta(\frac{1}{\sqrt{n \log n}}) \), where \( n \) is the number of users in the network. This shows diminishing throughput per user as the scale of the network becomes larger. Moreover, multi-hop forwarding suffers from large latency as the number of hops increases. These issues constrain the popularity of ad-hoc network in practical use.

We aim to exploit the benefits of multi-hop cooperation while avoiding its intrinsic issues by studying it from a different angle. We are mostly interested in the hybrid network, where the ad-hoc network performs as an auxiliary to an infrastructure-based network, such as cellular network, to extend the coverage, offload traffic volume and increase throughput. It is proved in [19, 20] that the hybrid network achieves close to \( \Theta(1) \) per user throughput. However, for practical system designs based on hybrid networks, there still exist several challenges:

- **How to design practical protocols to enable the transmissions over a hybrid network?**

- **How much application layer performance gains can the hybrid network bring?**

We aim to answer the above questions by studying the video multicast service over a hybrid cellular/ad-hoc network. Video streaming over wireless networks is known as a challenging task due to its stringent requirements for bandwidth and delay. We propose a novel solution that efficiently integrates scalable video coding, 3G broadcast and ad-hoc forwarding to maximize the overall video quality perceived by all viewers in the multicast...
session. Video layers are transmitted with different coverage and the users receiving more video layers can forward the extra layers to the users on the boundary, thus improve their performance through cooperation. Based on this idea, we have designed practical helper discovery and routing protocols to efficiently distribute the video content. Optimal resource allocation is conducted on the base station to maximize the total utility of all the users. With extensive trace driven simulations, we show that ad-hoc forwarding along with scalable video multicast significantly improves the users’ quality of experience.

1.1.2 PHY Layer Cooperative Communications

The previous multi-hop cooperative networking relies on the data forwarding of the intermediate users to enable end-to-end communication. Each hop is still point-to-point in the physical layer and cooperation only happens in the MAC layer and above. However, a point-to-point link is always unreliable due to interference, path-loss and user mobility. On the other hand, as a wireless channel is by nature a broadcast one, when one user transmits, the users around it will overhear partial or even all the information. Cooperative communication is a new technique in the PHY layer to utilize the surrounding users as relays to improve the reliability of the point-to-point link.

Figure 1.2 shows an illustrative example of the communication between a source and destination pair. A set of relays that are able to overhear the source signal can process and re-transmit them. The destination then combines the signals coming from the source and the relays. The first phase from source to the relays is called Source Broadcast Phase, while the second phase from the relays to the destination is called Cooperative Phase. Two main benefits can be exploited from such cooperation. First, transmissions of the same signal from multiple relays create spatial diversity and improve robustness against channel variations. Second, a single low data rate point-to-point link is replaced by two high speed links, which potentially improves the throughput of the transmission.

PHY layer cooperative communication techniques were intensively studied in [21–24]. Specifically, single relay cooperation schemes are presented in [21] and distributed space-time code (DSTC) based multi-relay cooperation is studied in [23, 24]. MAC layer
innovations aiming to recruit relays and coordinate transmission and reception is also necessary to be integrated with PHY layer advances. CoopMAC [25] is one of the first MAC layer design enabling cooperation under the IEEE 802.11 standards. As CoopMAC only recruits one relay at a time, a cooperative MAC layer incorporating DSTC and use multiple relays is proposed later on [26]. Due to the promising benefits achieved by the cooperative communication, WiMAX and LTE groups have also incorporated the use of relay stations in their standards.

As a general PHY layer technique, cooperative communications can be employed in many wireless communication systems to enhance the performance. Cognitive radio is an emerging technology to enable the secondary users (SUs) to dynamically access the licensed spectrum held by the primary users (PUs) in order to increase the efficiency of spectrum utilization. After some pioneering work [27], it has received considerable attention from both industry and academia. The summary of recent research advances of cognitive radio can be found in [28–30]. The combination of cooperative communications and cognitive radio is envisaged to bring various benefits. For example, cognitive relay network enhances the performance of the secondary network [31]. Cooperative spectrum sensing improves the reliability of detecting active primary users [32]. Recently, a new paradigm termed Cooperative Cognitive Radio Networks (CCRNs) has been advocated [33]. In CCRN, PUs recruit some SUs to cooperatively relay the primary traffic, and

Figure 1.2: An example of cooperative communications
in return grant them a portion of the channel access time. Not only can the PUs enjoy a significant throughput gain due to the cooperation, the SUs obtain opportunities to transmit their own data, resulting in a “win-win” situation.

We mainly focus our study on CCRN and trying to answer the following questions:

- **How can we further improve the performance of both PUs and SUs in CCRN?**
- **How to model competition and cooperation in the context of CCRN?**

Specifically, we propose to utilize Multiple-Input Multiple-Output (MIMO) under the framework of CCRN. By using the spatial multiplexing enabled by MIMO, the SUs equipped with multiple antennas can cooperatively relay the traffic for the PUs, while concurrently accessing the same channel to transmit their own traffic. Based on this idea, we have designed the beamforming vectors and analyzed the achievable link data rates. For the scenario of one primary link and multiple SUs, we formulate an optimization model based on a Stackelberg game to maximize their respective utilities. For the scenario of multiple PUs and SUs, we study the problem of finding an optimal matching between them. Extensive simulation results show that our scheme achieves larger “win-win” situation for both PUs and SUs.

### 1.1.3 Cooperation among Network Service Providers

We also investigate cooperation among network service providers. Generally speaking, the cooperation between two or more distinct organizations, such as companies, is driven by the market as well as their common interest. In the domain of wireless networks, network deployers or service providers may cooperate with each other, as long as it generates benefits to each of them. The study of such cooperation can better guide their network planning and service strategies.

However, to the best of our knowledge, little existing research work falls into this category. In [34], the authors study the cooperation between two co-located wireless sensor networks, where their nodes can forward the packets for each other and thus reducing the
energy consumption. A Similar idea is put into the scenario of wireless mesh networks in [35] and the authors present an analytical framework to identify the cooperation opportunities that are beneficial for both network providers. Figure 1.3 shows an example of the network topology discussed in [34, 35]. Further in [36], the authors propose the idea of sharing the spectrum between several wireless service providers.

We are specifically focused on the study of the cooperation among cellular operators. As cellular operators are devoting most of their efforts in expanding their respective networks due to the urgency of the data capacity crisis and market competition, it has become apparent that there exist huge variations in spectrum usage, channel quality and coverage in different operators’ networks [37]. Such diversity generates penalty of cooperation opportunities, which can be exploited to improve their network performance, spectrum efficiency and user experience. Moreover, all cellular infrastructure are converging to Long Term Evolution (LTE), which implies minimal modification to their infrastructure to enable their cooperation. We are interested in answering the following questions:
• **In what ways can the cellular operators cooperate?**

• **How much benefit can be generated from their cooperation?**

To answer the first question, we start from two simple cooperation strategies that we envisage the cellular operators may adopt. One is called FLEXROAM (short for “flexible roaming”), where the cellular operators allow their users to freely connect to a base station of either operator that provides the best signal strength. Another strategy called MERGER, where in addition to FLEXROAM, all the BSs of an operator can reuse the spectrum of its cooperating operators. This may be the result of a full merger of the two operators, but does not preclude a mutual business agreement short of a full merger.

To answer the second question under the FLEXROAM and MERGER strategies, we present an analytical model for indentifying the average user throughput gains using stochastic geometry. We assume all the BSs and users are distributed following independent Poisson processes and derived tractable closed-form equations. Further, we perform extensive simulations using real BS locations and OFDMA resource allocation algorithms. Both analytical modeling and simulations validate the advantage of operator cooperation in a realistic setting.

### 1.2 Dissertation Outline

This thesis is organized as follows.

In Chapter 2, we utilize multi-hop cooperative networking in the scenario of video multicast in cellular networks. We are interested in how such cooperation influences the application layer performance. Specifically, we propose a novel scalable video broadcast/multicast solution, called SV-BCMCS, that efficiently integrates scalable video coding, cellular broadcast and ad-hoc forwarding to maximize the overall video quality of all viewers in the 3G multicast session. We formulate and efficiently solve the optimal resource allocation problem of mapping each video layer to a modulation coding profile, and design the practical helper discovery and relay routing algorithms. Also, we analytically study the
gain of using multi-hop relaying. We then evaluate the designs with extensive trace driven simulations based on OPNET.

In Chapter 3, we focus on the physical layer cooperative communications and study how to best exploit its benefits in cooperative cognitive radio networks. By assuming the secondary users are equipped with multiple antennas, MIMO-based cooperation is investigated. We propose a novel MIMO-CCRN framework, which enables the SUs to cooperatively relay the traffic for the PUs while concurrently accessing the same channel to transmit their own traffic.

In Chapter 4, we consider cooperation between wireless network service providers. The benefits of sharing infrastructure and spectrum among multiple cellular operators are studied. We provide an analytical model based on stochastic geometry to identify the average user throughput under different sharing strategies, which gives tractable and accurate results. Moreover, we conducted extensive simulations for a multi-cell OFDMA system using real base station locations. The results of both analysis and Monte-Carlo simulations are compared and discussed.

Chapter 5 concludes the thesis and discusses some future research directions.
Chapter 2

Scalable Video Multicast in Hybrid Cellular/Ad-Hoc Networks

In this chapter, we study the benefits of cooperation to the video streaming performance of users in a cellular network. We propose a novel scalable video broadcast/multicast solution, SV-BCMCS, that efficiently integrates scalable video coding, cellular broadcast and multi-hop cooperation to balance the system-wide and worst-case video quality of all viewers in a cellular cell. We solve the optimal resource allocation problem in SV-BCMCS and develop practical helper discovery and relay routing algorithms. In addition, we analytically study the gain of using ad-hoc relay, in terms of users’ effective distance to the base station. Through extensive real video sequence driven simulations, we show that SV-BCMCS significantly improves the system-wide perceived video quality. The users’ average PSNR increases by as much as 1.70 dB with slight quality degradation for the few users close to the cell boundary. The main technical content of this chapter was presented in our previous works [38–40].

2.1 Background and Motivation

User demands for content-rich multimedia are driving much of the innovation in wire-line and wireless networks. Watching a movie or a live TV show on their cell phones, at
anytime and at any place, is an attractive application to many users. Mobile video broadcasting service, or mobile TV, is expected to become a popular application for cellular network operators. The service is currently operational, mainly in the unicast mode, with individual viewers assigned to dedicated radio channels. However, a unicast-based solution is not scalable. A Broadcast/Multicast service over cellular networks is a more efficient solution with the benefits of low infrastructure cost, simplicity in integration with existing voice/data services, and strong interactivity support. Thus, it is a significant part of cellular service. For instance, Broadcast/Multicast Services (BCMCS) [41,42] has become a part of the standards in the Third Generation Partnership Project 2 (3GPP2) [43] for providing broadcast/multicast service in the CDMA2000 setting. However, most existing BCMCS solutions employ a single transmission rate to cover all viewers, regardless of their locations in the cell. Such a design is sub-optimal. Viewers close to the base-station are significantly “slowed down” by viewers close to the cell boundary. The system-wide perceived video quality is far from reaching the social-optimum.

To improve the video multicast performance of cellular network, we utilize multi-hop cooperation among users. This is feasible as there is an ever larger number of phones equipped with multiple air interfaces, such as WiFi and Bluetooth. Using ad-hoc links to help data transmissions in cellular networks has been studied by several research groups in the past. In [44], a Unified Cellular and Ad-hoc Network (UCAN) architecture for enhancing cell throughput has been proposed, which is the starting point of our work. Clients with poor channel quality select clients with better channel quality as their proxies to receive data from the cellular base station. A packet to a client is first sent by the base station to its proxy node through a cellular downlink channel. The proxy node then forwards the packet to the client through an ad-hoc network composed of other mobile clients and IEEE 802.11 wireless links. In [45], the authors discovered the reason for the inefficiency that arises when using an ad-hoc peer-to-peer network as-is in a cellular system. Then they proposed two approaches to improve the performance, one of which is to leverage assistance from the base station, and the other is to leverage the relaying capability of multi-homed hosts. While the above articles are focused on the unicast data transmission
in cellular networks, ad-hoc transmission can be also employed to improve the performance of cellular Broadcast/Multicast Services. Based on UCAN, Park and Kasera [46] developed a new proxy discovery algorithm for cellular multicast receivers. The effect of ad-hoc path interference is taken into consideration. In ICAM [47], the authors developed a close-to-optimal algorithm for the construction of a multicast forest for an integrated cellular and ad-hoc network. Sinkar et al. [48] proposed a novel method to provide QoS support by using an ad-hoc assistant network to recover the loss of multicast data in the cellular network. However, the aforementioned works did not make use of scalable video coded streams, which allow cellular operators to flexibly select the operating point so as to strike the right balance between the system-wide aggregate performance and individual users’ perceived video quality.

Another technique to further improve performance is to employ Scalable Video Coding (SVC) [49]. SVC has been advocated to provide flexible video quality adjustment adapt to network and terminal capabilities [50]. By simply extracting the appropriate bit streams, it eliminates the computationally demanding transcoding processes at the video servers. It is shown that the latest SVC standard can achieve coding efficiency comparable to the state-of-the-art H.264/AVC non-scalable coding [51]. SVC can provide a wide range of temporal, spatial and amplitude scalabilities. Figure 2.1 illustrates the typical structure of temporal scalable video with B frames. The video is encoded into three layers with layer 1 containing all I and P frames, layer 2 containing all Bs frames, and layer 3 containing all B frames. Inter-frame prediction among the pictures follows a dyadic pattern. P frames are predicted from previous P-frames or I-frames. Bs frames are predicted from both I and P frames and from Bs frames. B frames are predicted from I and P frames and Bs frames, but not other B frames. Thus the higher layers cannot be decoded unless the lower layers are correctly decoded.

SVC combined with adaptive modulation and coding to maximize the video quality in an infrastructure based wireless network is studied in [52–54]. They differ in the formulation of the optimal resource allocation problem. Peilong et al. [55] and Donglin et al. [56] extended this idea into multi-carrier and cognitive radio scenarios. None of them, however,
We propose a novel scalable video broadcast/multicast solution, SV-BCMCS, that efficiently integrates scalable video coding, cellular broadcast and ad-hoc forwarding to achieve the optimal trade-off between the system-wide and worst-case video quality perceived by all viewers in the cell. In our solution, video is encoded into one base layer and multiple enhancement layers using SVC. Different layers are broadcast at different rates to cover viewers at different ranges. To provide basic video service to all viewers, the base layer is always broadcast to the entire cell. The channel resources/bandwidth of the enhancement layers are optimally allocated to maximize the system-wide video quality given the locations of the viewers and radio resources available at the base station. In addition, we allow viewers to forward enhancement layers to each other using short-hop and high-rate ad-hoc connections. Our work is also relevant to the studies of scalable video transmission over pure ad-hoc networks such as [57], where the authors consider the scenario of scalable video streaming from multiple sources to multiple users. A distributed scheme that optimizes rate-distortion is introduced. For our case, we study the scalable video streaming in cellular networks with the assistance of an ad-hoc network formed among users’ mobile devices. The optimization problem is solved at the cellular base station. In summary, our major contribution include:

1. We study the optimal resource allocation problem for scalable video multicast in cellular networks. We show that the system-wide video quality can be significantly increased by jointly assigning the channel resources for enhancement layers. Our
solution strikes a good balance between the average and worst-case performance for all viewers in the cell.

2. For ad-hoc video forwarding, we design an efficient helper discovery scheme for viewers to obtain additional enhancement layers from their ad-hoc neighbors a few hops away. Also a multi-hop relay routing scheme is designed to exploit the broadcast nature of ad-hoc transmissions and eliminate redundant video relays from helpers to their receivers.

3. We develop analytical models to study the expected gain of few-hop ad-hoc video relays under random user distribution in a cellular cell. An ad-hoc relay shortens the effective distance of the user to the base station, which is crucial to the throughput improvement of wireless transmission. Based on the models, we give the mathematical solution for the effective distance gain. The analysis underpins our protocol design from a theoretical viewpoint.

4. We selected three representative video sequences, and conducted video trace driven simulations using OPNET. We systematically evaluated the performance improvement attributed to several important factors, including node density, the number of relay hops, user mobility and coding rates of video layers. Simulations show that SV-BCMCS significantly improves the video quality perceived by users in practical cellular/ad-hoc hybrid networks.

### 2.2 Scalable Video Broadcast/Multicast Service (SV-BCMCS)

#### 2.2.1 System Architecture

In SV-BCMCS, through SVC coding, video is encoded into one base layer and multiple enhancement layers. Viewers who receive the base layer can view the video with the minimum quality. The video quality improves as the number of received layers increases. An enhancement layer can be decoded if and only if all enhancement layers below
Figure 2.2: Architecture of SV-BCMCS over a hybrid network (assuming three layers of video content).

It are received. The multicast radio channel of the base station is divided into multiple sub-channels. Here a sub-channel stands for either a temporal or frequency division of the bandwidth. Different layers of video are broadcast using different sub-channels with different coverage ranges. To maintain the minimum quality for all viewers, the base layer is always broadcast using a sub-channel to cover the entire cell. To address the decoding dependency of upper layers on lower layers, the broadcast range of lower layers cannot be shorter than that of the higher layers.

Figure 2.2 depicts the system architecture of SV-BCMCS as with eleven multicast users and three layers of SVC video, denoted by $L_1$, $L_2$, and $L_3$. The base station broadcasts three layers using three sub-channels with their respective coverage areas shown in the figure. All the users receive $L_1$ from the base station directly, while four users receive $L_2$ and two users receive $L_2$ and $L_3$ as well. With an ad-hoc network, the coverage of enhancement layers is extended further. As an example, user $a$ is in the coverage of $L_3$ and user $b$ is in the coverage of $L_2$. User $a$ relays $L_3$ to user $b$, who then relays $L_3$ to users $c$ and $d$. Meanwhile, user $b$ relays $L_2$ to users $c$ and $d$. Effectively, all four users $a$, $b$, $c$, and $d$ receive all three layers through the combinations of base station broadcast and ad-hoc relays. The key design questions of the SV-BCMCS architecture are:
Table 2.1: Notations

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r_i$</td>
<td>selected transmission rate (PHY mode) for layer $i$ video content</td>
</tr>
<tr>
<td>$p_i$</td>
<td>allocated fraction of channel for transmission of layer $i$ video content in cellular system</td>
</tr>
<tr>
<td>$n_i$</td>
<td>number of multicast users that can receive layer $i$ video/content in cellular domain</td>
</tr>
<tr>
<td>$R_i$</td>
<td>encoding rate of an individual layer $i$.</td>
</tr>
<tr>
<td>$L$</td>
<td>total number of layers</td>
</tr>
<tr>
<td>$N$</td>
<td>total number of multicast users in the cell domain</td>
</tr>
<tr>
<td>$\mathcal{N}$</td>
<td>the set of multicast users in the cell domain. $</td>
</tr>
<tr>
<td>$\Phi$</td>
<td>set of all possible transmission rates (or PHY modes)</td>
</tr>
<tr>
<td>$U_{total}$</td>
<td>aggregate utility of users in the cell domain</td>
</tr>
<tr>
<td>$d$</td>
<td>distance of the user to the base station</td>
</tr>
<tr>
<td>$r_t$</td>
<td>ad-hoc transmission radius/range for each user</td>
</tr>
<tr>
<td>$G$</td>
<td>effective distance gain (a random variable)</td>
</tr>
<tr>
<td>$S_e$</td>
<td>total area of the entire cell</td>
</tr>
<tr>
<td>$D$</td>
<td>the radius of the cell</td>
</tr>
</tbody>
</table>

1. How to allocate the radio resources among sub-channels to different video layers to strike the right balance between system-wide and worst-case video quality among all users?

2. How to design an efficient helper discovery and relay routing protocol to maximize the gain of ad-hoc video forwarding?

We examine these questions through analysis and simulations in the following sections. The key notation used in this chapter are shown in Table 2.1

### 2.2.2 Optimal Resource Allocation in Layered Video Multicast

Our objective for radio resource allocation is to maximize the aggregate user perceived video quality while providing a base-line minimum quality service for all users. The perceived video quality can be measured by PSNR (Peak Signal-to-Noise Ratio) or distortion, with $\text{PSNR} = 10 \log_{10} \left( \frac{M^2}{D} \right)$ ($D$ is the distortion represented by the MSE (Mean Square Error)).
Error) between the original image and the reconstructed image, and where $M_I$ is the maximum pixel value, typically set to be 255). The PSNR or distortion of a video sequence is the average of the corresponding measurements over all images in the same video sequence.

Modeling the distortion or the PSNR as the function of the user’s received data rate is an ongoing research topic. For example, in [58], distortion is modeled as a continuous function of video rate. In [57], the distortion with SVC is modeled as discrete values depending on the number of layers received by the user. In [59] and [60], PSNR is modeled as a linear/piece-wise linear function of the video rate with SVC. Here we employ a general non-decreasing utility function $U(R_{rec})$ in the optimization formulation, with $R_{rec}$ being a user’s receiving data rate. In Section 2.5, we replace the general utility function by a video sequence’s actual PSNR values.

Assume there are $L$ video layers, and the video rate of each layer is a constant $R_i$, $1 \leq i \leq L$. The broadcast channel is divided into $L$ sub-channels through time-division multiplexing. Each sub-channel can operate at one of the available BCMCS PHY modes. Each PHY mode has a constant data transmission rate and a corresponding coverage range.

Layer $i$ is transmitted using sub-channel $i$. Let $p_i$ be the time fraction allocated to sub-channel $i$, and $r_i$ be the actual transmission rate, or the PHY mode, employed by sub-channel $i$. In practice, to support the video rate $R_i$ of layer $i$, $r_i \cdot p_i \geq R_i$. We let $p_i = \frac{R_i}{r_i}$ in our formulation. In addition, the summation of the time fractions must be less than one $\sum_{i=1}^{L} p_i \leq 1$.

Suppose $n_i$ multicast users can receive the $i$-th layer video. Therefore $n_i - n_{i-1}$ users, $1 \leq i < L$, receive $i$ layers of video, while $n_L$ users receive all $L$ layers. The aggregate utility for all users is:

$$U_{total} = \sum_{i \in \mathcal{N}} U(R_{rec}^i)$$

$$= (n_1 - n_2) \cdot U(R_1) + \cdots + (n_j - n_{j+1}) \cdot U(\sum_{i=1}^{j} R_i) + \cdots + n_L \cdot U(\sum_{i=1}^{L} R_i)$$

$$= \sum_{i=1}^{L} (n_i - n_{i+1}) U(\sum_{j=1}^{i} R_j).$$

(2.1)
with \( n_{L+1} = 0 \). For a fixed total number of multicast users \( N \) in the cell domain, maximizing the average utility of multicast users is the same as maximizing the aggregate utility \( U_{total} \) in Eqn. (3.1).

Next, the computation of \( n_i \) is discussed. Let us first consider the base station transmission part. The number of users that can receive layer \( i \) directly from the base station is determined by the transmission rate \( r_i \) for sub-channel \( i \) (specifically, determined by the receiving SNR for the PHY mode with rate \( r_i \)). We can define this number of users as \( f_{BS}(r_i) \) for layer \( i \). Due to path loss, fading, and user mobility, \( f_{BS}(r_i) \) varies with \( r_i \) and generally is a monotonically decreasing function of \( r_i \). Thus the higher the transmission rate of the base station, the fewer the multicast users that can achieve the receiving SNR requirement and thus correctly receive the data.

In the second step, we take the ad-hoc relay into consideration. In SV-BCMCS, users who cannot receive layer \( i \) from the base station directly may receive it through the ad-hoc network. We use \( f_{AD-HOC}(r_i) \) to denote the number of users obtaining the layer \( i \) video through ad-hoc relay from other users. Thus \( n_i = f(r_i) = f_{BS}(r_i) + f_{AD-HOC}(r_i) \). In the system design, the base station gathers information about which user is getting data from which helper to compute \( n_i \). This is done in the helper discovery protocol illustrated in Section 2.3.

The optimal radio resource allocation problem can be formulated as the following utility maximization problem:

\[
\max_{\{r_i\}} \quad U_{total} = \sum_{i=1}^{L} [f(r_i) - f(r_{i+1})]U(\sum_{j=1}^{i} R_j),
\]

subject to:

\[
\begin{align*}
& r_i \leq r_j & i \leq j \text{ and } i,j = 1,2,\ldots,L, \\
& \sum_{i=1}^{L} \left( \frac{R_i}{r_i} \right) \leq 1, \\
& f(r_1) = N, \\
& r_i \in \Phi.
\end{align*}
\]
The objective is to find a set of transmission rates \( r_i \) for individual layers so as to maximize the aggregate utility. The constraint given by (2.3) ensures that the coverage of lower layers is larger than that of the higher layers. Constraint (2.4) guarantees the sum of sub-channels is no greater than the original channel. Constraint (2.5) ensures that the base layer covers the whole cell to provide basic video service to all the users. Finally, \( \Phi \) is the set of possible transmission rates (or PHY modes). Note that the traditional broadcast/multicast with one single stream is a special case of the above optimization problem with \( L = 1 \).

We have also designed a dynamic programming algorithm to solve the optimization problem formulated above, which works for any general non-decreasing function \( U(\cdot) \). The details are illustrated next.

### 2.2.3 Dynamic Programming Algorithm Solution

The optimization problem formulated above can be solved by a dynamic programming algorithm. To facilitate the design of the dynamic programming algorithm, a time unit of the original channel is divided into \( K \) equal length time slots. Sub-channel \( i \) broadcasts the \( i \)-th layer of rate \( R_i \), which requires \( \lceil \frac{R_i}{r_i} \cdot K \rceil \) time slots. It is desirable that \( \{ \frac{R_i}{r_i} \cdot K \} \) are integers so as to avoid the channel bandwidth wastage. This can be achieved by selecting \( K \) to be \( 10^n \) where the fraction numbers \( \{ \frac{R_i}{r_i} \} \) are represented using at most \( n \) significant digits.

The objective function shown in Eqn. (2.2) can be transformed as:

\[
U_{total} = \sum_{i=1}^{L} [f(r_i) - f(r_{i+1})]U(\sum_{j=1}^{i} R_j) = \sum_{i=1}^{L} f(r_i)[U(\sum_{j=1}^{i} R_j) - U(\sum_{j=1}^{i-1} R_j)],
\]

with \( U(\sum_{j=1}^{i-1} R_j) = 0 \) for \( i = 1 \). Transmitting the \( i \)th layer with rate \( r_i \) contributes a utility gain of \( f(r_i)[U(\sum_{j=1}^{i} R_j) - U(\sum_{j=1}^{i-1} R_j)] \) to the total utility \( U_{total} \). Define \( S_k^i \) to be the transmission rate of the \( i \)th layer that gives the maximal utility gain using no more than \( k \).
time slots, and $U_i^k$ to be the corresponding maximum utility gain. We have

\begin{align}
S_i^k &= \min\{r_i \in \Phi \cup \{\infty\} \mid \frac{R_i}{r_i} K \leq k\}, \\
U_i^k &= f(S_i^k) \cdot [U(\sum_{j=1}^{i} R_j) - U(\sum_{j=1}^{i-1} R_j)] \quad \forall 1 \leq i \leq L.
\end{align}

Since $f(r_i)$ is a non-increasing function, smaller $r_i$ leads to greater $f(r_i)$. Given the number of time slots $k$, $r_i$ should be the smallest PHY mode in $\Phi$ that can broadcast $i$-th layer. In case the value of $k$ is too small and no PHY mode in $\Phi$ can satisfy the condition of $\frac{R_i}{r_i} K \leq k$, the $i$-th layer cannot be broadcasted. The value of $S_i^k$ is set to be $\infty$. We let $f(\infty) = 0$ and the correspondingly $U_i^k$ to be zero.

We further define $\mathcal{U}_i^k$ to be the maximal utility gain of transmitting the first $i$ layers (from layer 1 to layer $i$) with the aggregate number of time slots no greater than $k$, and define $S^k_i$ to be the corresponding transmission rate vector for the first $i$ layers. Algorithm 1 with $\mathcal{U}_L^k$, $S^k_L$ giving the optimal solution.

In the above algorithm, line 5 to line 11 updates $U_i^k$ to include a new layer at each iteration. Line 7 and 8 solves the maximal utility problem of $\mathcal{U}_i^k$. Line 9 expands the transmitting rate vector $S^k_i$ by appending the optimal transmitting rate of layer $i$. Condition $S_{i-1}^m [i-1] \leq S^{k-m}_i$ in line 7 ensures that the broadcasting ranges of the higher layers are no greater than the lower layers, satisfying the Constraint (2.3). Note the algorithm is for a general problem without considering Constraint (2.5). If Constraint (2.5) is in effect, we
can replace the $K$ in line 6 with $K' = K\left(1 - \frac{R_1}{r_1}\right)$, where $r_1$ is the highest rate in $\Phi$ that can cover the entire cell (highest rate of $r_1$ uses the least number of time slots). Then we solve the optimization problem with $L - 1$ layers, i.e., from layers 2 to $L$. The maximum utility is $f(r_1) \cdot U(R_1) + U_{\mathcal{L}}^{K'}$. The complexity of the algorithm is $O(LK^2)$.

### 2.3 SV-BCMCS: Protocol Design

In a pure SV-BCMCS solution, users closer to the base station will receive more enhancement layers from the base station. They can forward those layers to users further away from the base station through ad-hoc links. Ad-hoc video relays are done in two steps: 1) each user finds a helper in its ad-hoc neighborhood to download additional enhancement layers; 2) helpers merge download requests from their clients and forward enhancement layers through local broadcast.

#### 2.3.1 Greedy Helper Discovery Protocol

We design a greedy protocol for users to find helpers. A greedy helper discovery protocol in the cellular and ad-hoc hybrid network was first presented in [44]. In that paper every node of the multicast group maintains a list of its neighbors, containing their IDs and the average cellular downlink data rates within a time window. Users periodically broadcast their IDs and downlink data rates to their neighbors. Each user greedily selects a neighbor with the highest downlink rate as its helper. Whenever a node wants to download data from the base station, it initiates helper discovery by unicasting a request message to its helper. Then the helper will forward this message to its own helper, so on and so forth, until the ad-hoc hop limit is reached or a node with the local maximum data rate is found. The ad-hoc hop limit is set by a parameter called Time-To-Live (TTL). The base station will send the data to the last-hop helper. The helpers will forward the data in the reverse direction of helper discovery to the requesting node.

We employ a similar greedy helper discovery mechanism. But unlike the case considered in [44], the locations and average cellular downlink data rates of helpers will affect the
resource allocation strategy of the base station in SV-BCMCS. In our scheme, the last-hop node in the path sends the final request message to the base station. Upon receiving this message, the base station updates the cellular data rate information of all the nodes along the path as the last-hop node’s cellular rate, assuming the ad-hoc link throughput is much larger than the cellular data rate. After that, the base station might resolve the coverage function as \( f(r_i) = \{\text{Number of nodes with updated cellular rate above } r_i\} \). The optimal broadcast strategy can then be calculated by solving the optimization problem defined in (2.2).

Moreover, to facilitate efficient relay routing, a node also needs to keep information about the relay requests routed through itself. The last-hop helper in the path also sends the final request message back to the initiating node. Every user along this ad-hoc path will get a copy of this final message.

As an example, the whole process is shown in Figure 2.3. The numbers in parenthesis are the average cellular downlink rates for each node. The dashed line indicates the ad-hoc neighborhood. The straight solid line is the ad-hoc path with the arrow pointing to the helper. User \( C \) attempts to find a helper within two hops to improve its video quality. Its request goes through \( B \) to \( A \). \( E \) and \( F \) are ignored by \( C \) and \( B \), since they are not the neighbor with highest cellular rate. To this end, User \( A \) knows where user \( C \) is located by the reverse route of the path that user \( C \) followed to find user \( A \). User \( A \) sends the request message to the base station to indicate that user \( A \) will act as user \( C \)’s helper using the relay path along user \( B \). Meanwhile, User \( A \) also sends this message (confirmation message) back to user \( C \) confirming that user \( A \) will act as its helper.

Note that [44] also proposed another helper discovery protocol using flooding. Instead of unicast, each node broadcasts the request message hop by hop. This method enables a node to find the helper with the global maximum data rate within the ad-hoc hop limit range. However, considering the large overhead of flooding messages in the ad-hoc network, we only adopt the greedy helper discovery protocol within the SV-BCMCS context.
2.3.2 SV-BCMCS Relay Routing Protocol

The SV-BCMCS routing protocol runs after the greedy helper selection protocol and the optimal radio resource allocation. Assuming optimal radio resource allocation has been performed, the base station decides to transmit the $L$ layers with different rates $r_1, r_2, \cdots, r_L$. It will broadcast this information to every node in the cell. Moreover, in the greedy helper discovery phase, each node obtains the information for all the relay paths to which it belongs. The major goal of the relay routing protocol is to maximally exploit the broadcast nature of ad-hoc transmissions and merge multiple relay requests for the same layer on a common helper.

Essentially, each helper needs to locally determine which received layers will be forwarded to its requesting neighbors. For each node $n$, define the set $K = \{\text{all neighbors that use } n \text{ as one-hop helper}\}$, the forwarding decision will be calculated in a distributed manner as shown in Algorithm 2.

For the receiving part, each node receives packets that satisfy two conditions: (i) the packets are sent from its direct one-hop helper; (ii) the packets belong to a layer that the node cannot directly receive from the base station. Otherwise the node will discard the packets. That is, the node has no use for packets that are from the layer to which the node
belongs, or from a lower layer than the layer to which it belongs.

A relay routing example is illustrated in Figure 2.3. The dashed-dotted lines represent the coverage area of each layer. 27, 18, 12 and 5 are the physical transmission rates for layers 4, 3, 2 and 1. Suppose $L = 4$ and maximal hop number (TTL) is 2. For node B in the figure, nodes C and D use it as a direct one-hop helper. For D, $l_D = 2$ (it is the layer to which node D belongs) according to the figure, and within 2 hops, D’s highest expected layer is $L_D = 4$. The highest expected layer is the highest layer which a node can expect to receive through its helpers while constrained by the TTL. In the same way, we can derive, $l_C = 1$ and $L_C = 4$. Thus, for node B, $l_{\text{min}} = \min \{l_C, l_D\} = 1$ and $L_{\text{max}} = \max \{L_C, L_D\} = 4$. Therefore, node B will broadcast the packets in layers 2, 3 and 4. Since node D is in layer 2, it will receive packets from node B in layers 3 and 4 only. Meanwhile, node C will receive all the packets in layers 2, 3 and 4.

Note a local broadcast doesn’t use RTS/CTS exchange in a practical implementation based on IEEE 802.11. Instead, the nodes’ carrier sensing threshold can be set to a reasonable small value. In this way, we can reduce the number of collisions due to the hidden terminal problem while still keep a favorable spatial reuse factor in the network.

In practice, wireless channels are error-prone and link quality changes over time due to fading and interference. This poses a challenge particularly to the video transmission. Due to the use of spatial temporal prediction, a compressed video is susceptible to transmission

---

**Algorithm 2 Forwarding Algorithm in SV-BCMCS Routing Protocol for Node $n$**

1. $\mathcal{K} = \{\text{all neighbors that use } n \text{ as one-hop helper} \}$
2. **for** $k \in \mathcal{K}$ **do**
   3. Find the highest layer $l_k$ that $k$ can directly receive from the base station
   4. Find the highest layer $L_k$ that $k$ can expect from any potential helper
5. **end for**
6. $l_{\text{min}} = \min \{l_k, k \in \mathcal{K}\}$
7. $L_{\text{max}} = \max \{L_k, k \in \mathcal{K}\}$
8. node $n$ broadcasts the packets between layer $l_{\text{min}} + 1$ to $L_{\text{max}}$ to its one-hop neighbors.
errors. To overcome such problem, an appropriate error protection mechanism is necessary in the practical implementation. FEC (Forward Error Correction) is widely used as an effective means to combat packet losses over wireless channel [57,61,62]. Our relay routing protocol can be easily extended to support FEC by letting the helper nodes decode and regenerate the parity packets for each video layer before forwarding them. We will evaluate the performance of our system with FEC in Section 2.5.

2.4 Analysis of the Gain of Ad-hoc Relay

In this section, we analytically study the expected gain from using ad-hoc relays under a random node distribution in the cell. Through ad-hoc video relays, users receiving fewer layers of packets (users at the coverage edge/boundary) are able to obtain video/content layers that they otherwise would not or could not receive. From the base station’s viewpoint, an ad-hoc relay shortens a user’s *effective distance* to the base station. As shown in Figure 2.4, user A relays data to user B who then relays the data to user C. If we assume that the bandwidth of an ad-hoc link is much larger than the cellular multicast rate, both user B and C can be seen as located at the same place as user A, then have the same effective distance as user A.

We define the *distance gain* of a user as the difference between its original distance to the base station and its last helper’s distance to the base station. For example, the distance...
gain of user $C$ is the difference between $OC$ and $OA$, where $O$ denotes the position of the base station. We are particularly interested in this metric due to the fact that in wireless communications, the distance between the transmitter and the receiver fundamentally affects their transmission rate. In the following, we develop a probabilistic model to study the distance gain due to ad-hoc relays under a random node distribution. Note that such distance gain is an upper bound to the case with limited ad-hoc bandwidth, which is able to show the insight of benefits brought by ad-hoc relay.

The typical transmission rate of an ad-hoc network, such as a network using IEEE 802.11, is much larger than the rate of a cellular network. For instance, IEEE 802.11g supports data rates up to 54 Mb/s. We therefore ignore the effect of wireless interference in our analysis. The interference will be included in our OPNET based simulation in the next section. We also assume that the number of data relays, or relay hops, is small. Using only a small number of relays is more robust against user mobility, and reduces the video forwarding delay. Furthermore, a smaller number of relays also reduces the traffic volume in the whole ad-hoc network.

Let $G$ be the distance gain of an arbitrary user. Obviously, $G$ depends on the location of the user, as well as the locations of other users in the same cell. It is also a function of the ad-hoc transmission range, and the number of relay hops allowed. Denote by $f_G(\cdot)$ the probability density function (pdf) of $G$. We develop a model to characterize $f_G(\cdot)$ by assuming the users are uniformly distributed in the cell. Our approach, however, also applies to other distributions within the cell. The list of key notation is included in Table 2.1. Figure 2.5 depicts an arbitrary user and is used to study the user’s distance gain in the case of a one-hop and two-hop relay.

### 2.4.1 Distance Gain Using One-Hop Relay

Assuming that the user is $d$ distance away from the base station, and the ad-hoc transmission radius/range is $r_t$. All other users falling into the transmission range of the user are potential one-hop helpers. Following the greedy helper discovery protocol, the user closest to the base station is chosen as the relay node. To calculate $f_G(g)$, we need to calculate the
Figure 2.5: The one-hop (left) and two-hop (right) ad-hoc relay analysis for an arbitrary user.

As illustrated in the left part of Figure 2.5, the whole cell space is divided into three regions: $S_1$, $S_2$, and $S_3$. Since the relay node is closer to the base station than any other node falling into the transmission range of the user, there should be no node in $S_2$ in the left part of Figure 2.5. To achieve a distance gain of $[g, g + \Delta g]$, there should be at least one node falling into $S_3$. Since the area of $S_3$ is proportional to $\Delta g$, the probability that two or more nodes fall into $S_3$ is a higher order of $\Delta g$, and thus will be ignored. Therefore, the probability of the distance gain in the range of $[g, g + \Delta g]$ is the probability that when we randomly drop $N - 1$ nodes (excluding the user under study) into the cell, one node falls into $S_3$, no node falls into $S_2$ and $N - 2$ nodes fall into the remaining area $S_1$. Based on the multinomial distribution:

$$f_G(g) = \lim_{\Delta g \to 0} \frac{\Pr(N - 2 \text{ nodes in } S_1, \text{ no node in } S_2, \text{ one node in } S_3)}{\Delta g},$$

$$= \lim_{\Delta g \to 0} \frac{(N - 1)!}{(N - 2)!1!1!} (q_1)^{N-2} (q_2)^0 q_3^3 \frac{\Delta g}{\Delta g} = \lim_{\Delta g \to 0} (N - 1)q_1^{N-2} q_3^3 \Delta g. \quad (2.10)$$

where $q_1, q_2, q_3$ are the probabilities of users located in the area $S_1, S_2, S_3$. Due to the uniform distribution of the users, $q_i = \frac{S_i}{S_c}, i = 1, 2, 3$, where $S_c$ is the area of entire cell.

$S_2$ is the overlapping area of two circles. For two circles with a known distance $d_{12}$ between their centers and with radius of each circle $c_1$ and $c_2$, let $S_{II}(d_{12}, c_1, c_2)$ represent their overlapping area. The detailed derivation of $S_{II}(d_{12}, c_1, c_2)$ is available in [63]. For
our case, $S_2 = S_{II}(d, r_t, d - g)$. For a fixed $r_t$, $S_2$ is a function of $g$ and $d$, so we use $S_2(g, d)$ from this point on. It is easy to verify that

$$\lim_{\Delta g \to 0} q_1 = 1 - \frac{S_2(g, d)}{S_c} = 1 - \frac{S_{II}(d, r_t, d - g)}{S_c},$$  \hfill (2.11)

$$\lim_{\Delta g \to 0} \frac{q_3}{\Delta g} \cdot S_C = \frac{S_3}{\Delta g} = -\frac{dS_2(g, d)}{dg} = \frac{dS_{II}(d, r_t, d - g)}{dx} \bigg|_{x=d-g}. \hfill (2.12)$$

Consequently, we have

$$f_G(g, d) = \frac{N - 1}{S_c} \cdot \left(1 - \frac{S_{II}(d, r_t, d - g)}{S_c}\right)^{N-2} \cdot \frac{dS_{II}(d, r_t, x)}{dx} \bigg|_{x=d-g}. \hfill (2.13)$$

The expected one-hop distance gain for a user at a distance $d$ from the base station can be derived as:

$$\mu_G(d) = \int_0^{r_t} g f_G(g, d) \, dg. \hfill (2.14)$$

### 2.4.2 Distance Gain Using Two-Hop Relay

The way to derive the distance gain in the two-hop case is similar to the one-hop case. However the two hop case is computationally more complex. To accurately characterize the two-hop relay gain, we need to calculate the joint density of the distance gain of the first and second relay hop. As illustrated in the right part of Figure 2.5, let $g_1$ be the distance gain of the first-hop relay, $\theta$ be the angle between the first-hop helper and the user with the base station as the origin, $g_2$ be the distance gain of the second-hop relay. In a manner similar to the one-hop case, we can calculate the joint density function $f(g_1, \theta, g_2)$ by calculating the multinomial distribution of $N - 1$ nodes fall into five regions as illustrated in the right part of Figure 2.5. Let $n_i$ be the number of nodes falling into area $S_i$. We need to calculate the multinomial probability of $n_1 = N - 3, n_2 = n_4 = 0, n_3 = n_5 = 1$. The areas of $S_{1-5}$ are functions of $g_1, g_2, d$ and $\theta$.

The pdf of the joint distance gain can be calculated as:

$$f(g_1, g_2, \theta) = \lim_{\Delta g_1, \Delta \theta, \Delta g_2 \to 0} \frac{(N - 1)(N - 2)}{S_c^{N-1}} \cdot S_1^{N-3} \cdot \frac{S_3}{\Delta g_2} \cdot \frac{S_5}{\Delta g_1 \Delta \theta}. \hfill (2.15)$$

It can be determined that $S_5 = (d - g_1)\Delta \theta \Delta g_1$ and $S_1 = S_c - \sum_{i=2}^5 S_i$. Unfortunately, the calculation of $S_2, S_3$ and $S_4$ is fairly involved.
Figure 2.6: Reference figure for positions of the user, helper 1 and helper 2. We label the circles as $C_A, C_B, C_C$ and $C_D$, with $C_A$ and $C_B$ centered at the user and helper 1, $C_C$ and $C_D$ centered at the Base Station. I, II and III indicate the three different patterns that $C_D$ intersect with $C_A$ and $C_B$.

Now let’s take a closer look at the positions of the user, helper 1 and helper 2 in Figure 2.6. Note that the center positions and the radii of the circles $C_A, C_B, C_C$ and $C_D$ can be computed based on $g_1, g_2, d, r_1$ and $\theta$. We represent by $S_{II}(C_i, C_j)$ the overlapping area of two circles $C_i$ and $C_j$, $S_{III}(C_i, C_j, C_k)$ as the overlapping area of the three circles $C_i, C_j$ and $C_k$. $i, j$ and $k$ are chosen from $A, B, C, D$ in the example used herein. With known radii and center positions, the closed form formulas of $S_{II}(\cdot)$ and $S_{III}(\cdot)$ are derived in [63].

With different positions of the helper 2, $C_D$ may intersect with $C_A$ and $C_B$ in one of three patterns shown in Figure 2.6. Here the results are given directly:

**case I:**

$$S_2 = S_{II}(C_B, C_D), \quad \frac{dS_3}{dg_2} = -\frac{dS_2}{dg_2}, \quad S_4 = S_{II}(C_A, C_C) \quad (2.16)$$

In case II and III, $S_2$ and $S_4$ overlaps with each other and combine into one part, $S_{24}$.

**case II:**

$$\frac{dS_3}{dg_2} = -\frac{d[S_{II}(C_B, C_D) - S_{II}(C_A, C_D)]}{dg_2}$$

$$S_{24} = S_{II}(C_B, C_D) + S_{II}(C_A, C_C) - S_{II}(C_A, C_D) \quad (2.17)$$
case III:

\[
\frac{dS_3}{dg_2} = -\frac{d[S_{II}(C_B, C_D) - S_{III}(C_A, C_B, C_D)]}{dg_2} \\
S_{24} = S_{II}(C_B, C_D) + S_{II}(C_A, C_C) - S_{III}(C_A, C_B, C_D)
\]  

(2.18)

The joint pdf can now be calculated as

\[
f(g_1, g_2, \theta) = \left(\frac{N-1}{N-3}\right) \cdot \frac{S_{1}^{N-1}}{S_{c}^{N-1}} \cdot dS_3 \cdot (d - g_1),
\]

(2.19)

with different \(S_1\) and \(\frac{dS_3}{dg_2}\) as shown from (2.16) to (2.18). In case I, \(S_1 = S_c - S_2 - S_4\). And in case II and III, \(S_1 = S_c - S_{24}\). Finally, the expected two-hop distance gain for a user at the distance \(d\) can be calculated as

\[
\mu_{G_1+G_2}(d) = \int_0^{g_{2}^*} \Phi_{(III)}(g_1, g_2, d, \theta) dg_2 \\
+ \int_{g_2^{*}}^{g_2^*} \Phi_{(II)}(g_1, g_2, d, \theta) dg_2 + \int_{g_2^*}^{r_t} \Phi_{(I)}(g_1, g_2, d, \theta) dg_2
\]

(2.20)

with

\[
\Phi_{(i)}(g_1, g_2, d, \theta) = \int_{0}^{\theta^*} (g_1 + g_2) f_{G_1G_2(i)}(g_1, g_2, d, \theta) d\theta dg_1, \ i = I, II, III
\]

(2.21)

The critical values of \(g_2\) from case I to II and case II to III are defined as \(g_2^*\) and \(g_2^{**}\). From the law of cosines, we have:

\[
\theta^* = \arccos \frac{d^2 + (d - g_1)^2 - r^2}{2d(d - g_1)}
\]

(2.22)

2.4.3 Impact of ad-hoc wireless relay to user performance

With the concept of “distance gain”, we can think of more users moving closer to the base station and “appear to exist” within a certain distance of the base station compared to the scenario with no ad-hoc relay. The benefit it brings to our layered video multicast is that the video layers can be received by more users, which we have shown in the previous sections. Next we will analytically derive the number of users that effectively move closer to the base station with the aid of ad-hoc wireless relay.
Based on the model we build up, our objective is to calculate on average how many users outside a given distance \( d \) can move into the circle, with the aid of ad-hoc relay. If we suppose that different video layers are transmitted with different ranges, such an increase represents the number of additional users that can receive a certain video layer. So it has a significant practical meaning.

In detail, we divide the ring between a distance \( d \) and \( d + r_t \) into many concentric rings, each with a width of \( \Delta \). Note that \( r_t \) is the range of ad-hoc transmission. One-hop ad-hoc relay is considered in this case; however the approach can be applied to the multiple hop relay scenario. \( N \) is the total number of multicast users in the entire cell, and \( D \) is the radius of the cell. The average number of users in the \( k \)th ring is:

\[
N_k(\Delta) = N \cdot \frac{\pi[(d + k\Delta)^2 - (d + (k - 1)\Delta)^2]}{\pi D^2} = N \frac{\Delta[(2k - 1)\Delta + 2d]}{D^2}.
\] (2.23)

Then, the probability that a user in the \( k \)-th ring can move within distance \( d \) is:

\[
p_k(\Delta) = \int_{k\Delta}^{r_t} f_G(g, d) dg
\]

The average number of users that moves into the circle of radius \( d \) is:

\[
N_{\text{ave}} = \sum_{k=1}^{\left\lfloor \frac{D}{\Delta} \right\rfloor} N_k(\Delta) \cdot p_k(\Delta).
\] (2.24)

As \( \Delta \to 0 \), Equation (2.24) can be rewritten as:

\[
N_{\text{ave}}(d) = \int_0^{r_t} \frac{2N(d + r)}{D^2} \int_r^{r_t} f_G(g, d) dg dr.
\] (2.25)

Recall in our formulation (2.2), with the assistance of the ad-hoc network, the base station can reach a larger number of users \( f(r_i) = f_{BS}(r_i) + f_{AD-HOC}(r_i) \). Now \( f_{AD-HOC}(r_i) \) can be approximated by \( N_{\text{ave}}(d_i) \), where \( d_i \), the distance for certain transmission rate \( r_i \), is discussed and derived in section 2.2.2.

### 2.4.4 Numerical Results Using the Analytical Model

Based on the analytical model presented above, we numerically computed the resulting distance gain and user number increase. Since the pdf \( f_G(g, d) \) is a function of \( d \), the
The average distance gain (meters) for nodes with different distances to the base station.

User achieves different distance gain when its distance to the base station varies. The results in this section are for different node densities in a cell with a radius 1000 meters. We set the ad-hoc range at 100 meters. Therefore, if there are a totally 500 nodes, on average each node has $500 \cdot \frac{\pi 100^2}{\pi 1000^2} = 5$ neighbors.

Figure 2.7 shows how the distance gain $g$ derived in sections 2.4.1 and 2.4.2 changes with $d$ when TTL is set to one and two. For example, when the total number of multicast users is 700, for the user at the boundary of the cell, i.e., $d = 1000$, the expected one-hop and two-hop distance gains are $g_{TTL=1} = 62.86$ and $g_{TTL=2} = 121.61$.

For the same setting with 500 multicast users, we calculate the increase in the numbers of users at different ranges according to Equation (2.24). For $d = 500, 700$ and 900 meters, the “number increase/original number of users” are $28.57/125, 38.32/245, 48.08/405$ respectively. Note without ad-hoc, the original number of users is proportional to $d^2$ due to the uniform user distribution. We can observe an obvious increase as ad-hoc relay squeezes the users towards the base station. This explains the potential of video quality improvement by using ad-hoc network in the SV-BCMCS protocol.
2.5 Performance Evaluation

In this section, the performance of SV-BCMCS is evaluated using OPNET based simulations. Compared to the popular network simulators ns-2 and Omnet++ which are free, OPNET is a commercial network simulator using standards based models [64]. The performance of SV-BCMCS is compared with the performance of traditional 3G BCMCS under various scenarios. The impact of node density, node mobility, number of relay hops and the base layer video rate is investigated. Results demonstrate that SV-BCMCS consistently out-performs BCMCS with or without the aid of ad-hoc data relay.

2.5.1 Simulation Settings

Network Settings

SV-BCMCS is simulated using the wireless modules of OPNET modeler. It is assumed that all multicast users/nodes have two wireless interfaces: one supports a CDMA2000 channel for 3G video service, and the other supports IEEE 802.11g for ad-hoc data relay. The data rate of the ad-hoc network is set to be 54 Mb/s, and the transmission power covers 100 meters. Since OPNET Modeler does not provide built-in wireless modules with dual interfaces, the 3G downlink is simulated as if individual users generate their own 3G traffic according to the experimental data presented in [41, 65]. The free-space path loss model is adopted for 3G downlink channels, where the Path Loss Exponent (PLE) is set to 3.52, and the received thermal noise power is set to be -100.2dBm. Eleven PHY data rates are supported according to the 3GPP2 specifications [66].

The 3G cell is considered to be a circle with a radius of 1000 meters, with a base station located in the center. It is assumed that 3G BCMCS supports a physical layer rate of 204.8 kb/s, which is able to cover the entire cell using a (32, 28) Reed-Solomon error correction code, according to [41]. In our simulation, the transmission power of the 3G base station is set accordingly so that BCMCS can broadcast the video to the entire cell. The same base station transmission power is used in SV-BCMCS evaluations. The users'
Table 2.2: Rate (kb/s) and PSNR (dB) values of all the layers for the three SVC encoded video sequences

<table>
<thead>
<tr>
<th></th>
<th>Mobile</th>
<th>Football</th>
<th>Bus</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Rate PSNR</td>
<td>Rate PSNR</td>
<td>Rate PSNR</td>
</tr>
<tr>
<td>Base layer</td>
<td>149.1 31.1756</td>
<td>136.0 30.0436</td>
<td>138.5 31.2975</td>
</tr>
<tr>
<td>Enh. layer 1</td>
<td>237.5 31.9141</td>
<td>233.5 30.8987</td>
<td>217.3 31.9257</td>
</tr>
<tr>
<td>Enh. layer 2</td>
<td>320.2 32.8918</td>
<td>353.5 32.8803</td>
<td>319.8 33.3549</td>
</tr>
<tr>
<td>Enh. layer 3</td>
<td>391.8 33.9392</td>
<td>421.1 34.0621</td>
<td>422.0 35.3183</td>
</tr>
<tr>
<td>Enh. layer 4</td>
<td>461.1 35.4830</td>
<td>506.4 36.1888</td>
<td>494.3 37.4306</td>
</tr>
<tr>
<td>Enh. layer 5</td>
<td>522.0 37.1488</td>
<td>541.6 37.0958</td>
<td>545.1 39.1810</td>
</tr>
</tbody>
</table>

average PSNR is used as the metric of their received video quality.

**Scalable Video Settings**

Three standard SVC test video sequences, *Mobile*, *Football* and *Bus* in QCIF resolution (176 × 144 pixels), with a frame rate of 15 frames/sec are used in the simulations. All of the sequences are available from [67]. They are played repeatedly to yield video sequences with a length of approximately 90 seconds. Unless indicated otherwise, the video length is the simulation length for our simulation runs. We use JSVM 9.19.7 reference software to encode the video sequences into one base layer and five SVC fidelity enhancement layers with MGS (Medium-Grain fidelity Scalability), based on the SVC extensions of H.264/AVC [49]. In our setting, one GOP (Group-Of-Picture) includes 16 frames. By adjusting the quantization parameters (QP) for each layer in the encoding, all videos are encoded at the rate of about 530 kb/s. The resulting rate points and PSNR values for the layers of each encoded video sequences are summarized in the Table 2.2.

The generated packet information of each video sequence is integrated into OPNET Modeler to simulate video transmission and reception. Since we have modeled interference in OPNET, there will be dynamic packet loss in the ad-hoc network. Thus the video layers are encoded independently using (6, 5) FEC to combat the packet loss, as described
in Section 2.3. On the receiving side, a user device decodes an enhancement layer if it has successfully received enough parity packets of this layer, otherwise this layer and all the higher enhancement layers are discarded due to the decoding dependency. Finally, we use JSVM to decode the received stream for each user and measure the PSNR of the reconstructed video.

Note that although we can use any general non-decreasing utility function $U(\cdot)$, for simplicity, in our simulation we characterize the video stream by a set of PSNR points $u_l$, which represent the PSNR of the encoded video with $r_{l}^{enc} = \sum_{i=1}^{l} R_i$ being the corresponding encoding rate of the scalable video stream at layer $l$, $l \in 1 \ldots L$. Then we solved the optimal resource allocation problem in the simulation with this model. The $u_l$’s and $r_{l}^{enc}$’s are as listed in Table 2.2.

### 2.5.2 Stationary Scenarios

In stationary scenarios, a certain number of fixed nodes (users) are uniformly distributed in the 3G cell. The presented results are averages over ten random topologies. The 90% confidence interval is also determined for each simulation point.

#### The Impact of Node Density

In this scenario, we use the settings described in the previous section. The number of multicast receivers in the cell is ranging from 100 to 600, to simulate a sparse to dense node distribution. As a comparison, “Traditional BCMCS” or “BCMCS” indicate the transmission of a single layer video, which we encode using JSVM single-layer coding mode. We encode the single-layer video at approximately 185 kb/s (specifically, 180.6 kb/s for Mobile, 184.9 kb/s for Football and 191.3 kb/s for Bus). The rate is close to the maximum rate allowed by the BCMCS with the physical layer rate of 204.8 kb/s. The average PSNR for scenarios without an ad-hoc network, and scenarios with ad-hoc network (TTL set to 3) for all three video sequences, are given in Figure 2.8 and Figure 2.9.

With and without ad-hoc data relay, SV-BCMCS consistently out-performs the tradi-
tional BCMCS. Without ad-hoc relay, SV-BCMCS provides approximately 0.8 dB (Mobile sequence), 0.6 dB (Football sequence) and 0.2 dB (Bus sequence) gain, respectively. The ad-hoc relay leads to extra performance gain. For example, for video sequence Bus, when the number of multicast receivers is 100, the ad-hoc relaying gives a 0.36 dB additional PSNR improvement. When the receiver number is 600, the additional improvement reaches 1.17 dB. In general, the PSNR gain with ad-hoc relay increases as the number...
of users grows because more users facilitate ad-hoc relaying compared to the traditional BCMCS with 600 users, SV-BCMCS improves the users’ average PSNR by 1.70 dB for Mobile sequence, and 1.70 and 1.35 dB for Football and Bus sequences respectively.

The impact of number of relay hops (TTL)

Figure 2.10 depicts the performance of SV-BCMCS under different TTLs. The analysis in Section 2.4 studies the users’ effective distance gain with the aid of ad-hoc data relay. Here the impact of ad-hoc relay is examined in a more practical setting: in the presence of wireless interference and using real video sequences. The experiments are done with the number of users set to be 500.

As shown in the figure, the performance of SV-BCMCS improves as TTL increases. With more relaying hops, users can potentially connect to the helpers closer to the cellular BS, thus obtaining more higher enhancement layers via ad-hoc relaying. For example, in the figure of Bus sequence, with TTL = 1, the average PSNR is about 0.66 dB higher than that in the traditional BCMCS. Such an improvement increases to 1.54 dB when the TTL reaches four. A similar trend can be observed for the Mobile and Football sequences as
well. However, as the TTL becomes larger, the additional interference and communication overhead grow. Hence the PSNR curves flatten out gradually. Finally, note that even with TTL = 0, i.e., without ad-hoc relay, the SV-BCMCS still out-performs BCMCS due to the employment of SVC coding and the optimal resource allocation.

### 2.5.3 Mobile Scenarios

The impact of user mobility to the users’ received video quality, represented by PSNR, is studied next. The random walk model with reflection \([68]\) is used to drive the user movement in the simulation. The individual user’s moving speed is randomly selected in the range from zero to \(\text{maxspeed (m/s)}\), where \(\text{maxspeed}\) is a simulation parameter. Both moving speed and moving direction are adjusted periodically, with the time period drawn from a uniform distribution between zero and 100 seconds. We only present the results for the \(\text{Mobile}\) sequence, however similar observations are made for other video sequences. Also, we loop the video into a longer sequence with a length of 10 minutes. The mobility affects SV-BCMCS’s performance in the following two ways: (i) the efficiency of ad-hoc network degrades due to the link failures caused by mobility, and (ii) the optimal channel
allocation is disrupted since the user positions keep changing. SV-BCMCS periodically reconfigures the optimal allocation so as to adapt to the user position change.
Table 2.3: Impact of Reconfiguration Interval on the Average PSNR ($\text{maxspeed}=10 \text{ m/s}$)

<table>
<thead>
<tr>
<th>Reconfiguration Interval (sec)</th>
<th>45</th>
<th>30</th>
<th>10</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Users’ Average PSNR (dB)</td>
<td>32.498</td>
<td>32.569</td>
<td>32.697</td>
<td>32.783</td>
</tr>
</tbody>
</table>

**The Impact of moving speed**

There are 600 users in the cell, with ad-hoc relay hops (TTL) set to be three. The $\text{maxspeed}$ is set at 1, 2, 5, 10, 15 and 30 m/s, respectively. The base station reconfiguration interval of the optimal channel allocation is set to be 30 seconds. Figure 2.11 depicts the users’ average PSNR for different $\text{maxspeed}$. Clearly, the performance of the SV-BCMCS degrades as the users move faster, especially when the speed increases beyond 5 m/s. In fact, when the speed is above 5 m/s, the users’ average PSNR is approaching the value without the aid of ad-hoc data relays, as shown in Section 2.5.2.

**The Impact of the Reconfiguration Interval**

Table 2.3 summarizes the impact of optimal allocation reconfiguration interval on the average PSNR when $\text{maxspeed}$ is 10 m/s. As the interval becomes shorter, the SV-BCMCS can adapt to the mobile environment faster, leading to better performance. This is at the price of computational power, communication overhead, and changing video quality perceived by some users. Hence the reconfiguration interval should be selected to strike the right balance.

2.5.4 Tradeoffs between the Base Layer Rate and the Overall Performance

In SV-BCMCS, depending on a user’s location, he/she may receive the same video at different quality levels by receiving a different number of video layers. In the worst case, a user may only receive the base layer, which is transmitted to the entire cell. SV-BCMCS allows users having better channel condition to receive better quality video, which is fair in the sense of maximizing the aggregate utility for all users. The study of the right fairness
metric, however, is outside the scope of this dissertation. Here we focus on the tradeoffs between the base layer rate and the overall improvement of user perceived video quality.

Figure 2.12 depicts the average PSNR vs. the base layer rate in SV-BCMCS with and without ad-hoc data relay. The Mobile sequence is being used. There are 600 fixed users in the cell. We change the base layer rate by adjusting the QP for the base layer during encoding. The QP for the enhancement layer rate is fixed which gives a relatively similar encoding rate for the enhancement layers. The resulting rates for the base layer are 123.0, 136.1, 149.1 and 166.2 kb/s respectively. We can see that as the base layer rate increases further beyond 136.1 kb/s, the users’ average PSNR decreases. The base layer rate must not exceed the single layer rate (204.8 kb/s) as used in BCMCS. As we approach this rate, the entire channel can only transmit the base layer, and there is hardly any difference between SV-BCMCS and BCMCS. Intuitively, the high base rate leaves less “room” for SV-BCMCS to optimize and to achieve a higher average PSNR.

Figure 2.13 shows the users’ average PSNR vs. the distance to the base station in SV-BCMCS. The base layer rate is set to be 149.1 kb/s and 166.2 kb/s respectively and all the other settings are the same as above. Each point in the figure represents one user. Clearly, more users in the small base layer rate case (149.1 kb/s) are able to enjoy higher PSNR than in the large base layer rate case (166.2 kb/s). Specifically, 301 users in the first case are with PSNR more than 32.0 dB, compared with 208 users in the second case. Using less channel bandwidth to deliver a smaller rate base layer enables enhancement layers to be transmitted further. However, the users who only obtain the base layer perceive worse video quality in the small base layer rate case than in the large base layer rate case. With the aid of ad-hoc data relay, more users are able to receive video with higher quality regardless of the base layer rate.

2.6 Summary

In this chapter we present SV-BCMCS, a novel scalable video broadcast/multicast solution that efficiently integrates scalable video coding, cellular broadcast and ad-hoc
Figure 2.13: Tradeoffs between the base layer rate and the overall performance in SV-BCMCS.

forwarding. We formulate the resource allocation problem for scalable video multicast in a hybrid network whose optimal solution can be resolved by a dynamic programming algorithm. Efficient helper discovery and video forwarding schemes are designed for practical layered video/content dissemination through ad-hoc networks. Furthermore, we analyze the effective distance gain enabled by ad-hoc relay, which provides insight into the video quality improvement made possible by using ad-hoc data relay. Finally, OPNET based real
video simulations show that a practical SV-BCMCS increases the users’ average PSNR by 1.35 ∼ 1.70 dB for the video sequences we use, with the ad-hoc networks accounting for around 1.2 dB improvement. Moreover, SV-BCMCS still maintains a minimum of 0.80 dB performance improvement when the nodes’ moving speed is less than 5 m/s, while periodical reconfiguration is necessary in fast moving scenarios. The tradeoffs between the base layer rate and the overall performance is discussed and we demonstrate that SV-BCMCS can significantly improve the system-wide video quality, though a few viewers close to the boundary will suffer a slight quality degradation.
Chapter 3

Exploiting Multiple Antennas in Cooperative Cognitive Radio Networks

In this chapter, we discuss utilizing the PHY layer cooperative communications to benefit the cognitive radio networks. Specifically, we study how to take advantage of the Multiple-Input Multiple-Output (MIMO) technique in Cooperative Cognitive Radio Networks (CCRN), which is an unexplored area. We propose a novel MIMO-CCRN framework, which enables the secondary users (SUs) to utilize MIMO techniques to cooperatively relay the traffic for the primary users (PUs), while concurrently accessing the same channel to transmit their own traffic. Specifically, we consider two typical network scenarios. For the case of a general PU link and multiple SUs, we provide theoretical analysis for the link rates and then formulate an optimization model based on a Stackelberg game to maximize the utilities of PUs and SUs. In addition, we extend our analysis to a practical cellular network with multiple MIMO-empowered femtocells, and provide an algorithm to find a stable matching of the PUs and SUs. Evaluation results show that high utility gains are achieved by both PUs and SUs by leveraging MIMO spatial cooperation in our proposed framework. The technical content of this chapter appeared in our previous works [69, 70].

3.1 Background and Motivation

Cognitive radio, with the capability to flexibly adapt its transmission or reception parameters, has been proposed as the means for unlicensed secondary users (SUs) to dynam-
ically access the licensed spectrum held by primary users (PUs) in order to increase the efficiency of spectrum utilization.

Leveraging cooperative diversity to enhance the performance of cognitive radio networks has attracted much attention. One category of work focused on the cooperation between SUs. In [71], by exploiting the spectrum-rich but low traffic demand SUs to relay the data for other SUs, the overall performance of the secondary network can be improved. A relay-assisted routing protocol exploiting such spectrum heterogeneity was then proposed in [72]. Also, multiple SUs can cooperatively sense the spectrum to better detect the presence of the PUs and the corresponding work can be found in [28].

Recently, another category termed Cooperative Cognitive Radio Networks (CCRNs) has been advocated [33], which concentrates on cooperative communication between PUs and SUs. In CCRN, PUs recruit some SUs to cooperatively relay the primary traffic, and in return grant them a portion of the channel access time. Not only can the PUs enjoy a significant throughput gain due to the cooperation, the SUs obtain opportunities to transmit their own data, resulting in a “win-win” situation. Specifically in [33], the primary link may decide to lease the spectrum for a fraction of time to the SUs in exchange for their cooperation in relaying the primary data. This concept has been further extended to combine the pricing of the spectrum in [73], and to the multi-channel scenario in [74]. Recently, it was also studied in a dual infrastructure-based cognitive radio network with multiple primary links in [75]. In this chapter, we will focus our study on the design of new CCRN schemes.

Although the conventional CCRN framework benefits both the PUs and SUs, there still exist some inefficiencies. By assuming each device is equipped with a single antenna, previous work such as [33, 73, 75] only allows the channel sharing to happen in the temporal domain. In particular, as shown in Figure 3.1, a frame is time-divided into three phases. The first phase is used for the primary transmitter (PT) to broadcast the data to the relaying SUs. In the second phase, those SUs form a distributed antenna array to cooperatively relay the primary data to the primary receiver (PR). In return, the third phase is leased to the SUs for their own traffic. We can easily see from the above that, to motivate the SUs to cooperate,
Figure 3.1: An illustration of the conventional CCRN scheme.

The PU must completely give up its spectrum access to SUs for their transmissions in the third phase. This results in a large overhead to PU’s transmission and partially neutralizes the gain it achieves from cooperation. Additionally, the SUs’ transmissions are confined within the third phase, which is usually short. Considering there will be multiple secondary links competing for spectrum access, this phase will become crowded. As a consequence, the throughput gains obtained by either the PU or the SUs are still limited.

On the other hand, MIMO has been widely accepted as a key technology to increase wireless capacity. It is a physical layer technology that can provide many types of benefits through multiple antennas and advanced signal processing. Multiple independent data streams can be transmitted or received over the MIMO antenna elements via spatial multiplexing to improve channel capacity. Given its potential, MIMO has been adopted in next-generation WLAN and cellular network standards. Extensive research work on MIMO have been done at the physical layer for point-to-point and cellular communications [76]. Many researchers have exploited the benefits of MIMO from a cross-layer prospective. In wireless mesh networks, the throughput optimization problem based on MIMO was stud-
A MIMO-aware MAC and routing mechanisms are presented in [80, 81]. However, studies on MIMO in cognitive radio networks remain limited and mainly focus on the physical layer, as in [82, 83]. All existing work on CCRNs assumes single antennas and operate in the temporal domain only. They do not leverage the spatial domain in the cooperative transmission for the case when the nodes are equipped with multiple antennas. We consider this setting and seek to bridge this gap by providing a practical paradigm for taking advantage of the MIMO technique. To the best of our knowledge, this is the first work utilizing MIMO in the context of CCRNs.

To address the problems existing in conventional CCRNs, we propose a novel design called the MIMO-CCRN framework for cooperation among SUs and PUs by exploiting MIMO antennas on SUs’ transceivers. The basic idea of MIMO-CCRN can be explained using an example, as shown in Figure 3.2. A pair of PUs, typically legacy devices with a single antenna, are co-located with several SUs seeking transmission opportunities. The SU’s are equipped with multiple antennas. The PUs can improve their performance by recruiting SUs as relays. Assume SU$_2$ and SU$_3$ are selected as the relays. A time period is then divided into two phases. In Phase One, the primary transmitter broadcasts data to SU$_2$ and SU$_3$. Meanwhile, SU$_2$ can simultaneously receive its own traffic from another SU,
SU₁, as long as the total number of primary and secondary streams is no greater than its antenna Degree-of-Freedom (DoF) [84]. Similarly in Phase Two, SU₂ and SU₃ cooperatively forward the primary data to the primary receiver. At the same time, SU₃ is able to transmit its own data to SU₄ using beam-forming, if it ensures the interference from the secondary stream is cancelled at the primary receiver. As we can see, in the MIMO-CCRN framework, the SUs utilize the capability provided by MIMO to cooperatively relay the traffic for the PUs while concurrently obtaining opportunities to access the spectrum for their own traffic. The PU does not need to allocate a dedicated fraction of channel access time to SUs. Furthermore, the PU can still use legacy devices and is not required to change its hardware to support MIMO capability. MIMO-CCRN can greatly improve the performance of both PUs and SUs. Of course, the trade-off is that the SUs must be equipped with sophisticated MIMO antennas, which are expected to be widely adopted in future radio devices.

We focus on the cross-layer design and performance analysis of the proposed MIMO-CCRN framework. We are interested in answering the following questions: what are the benefits of exploiting MIMO in the context of CCRN; how the PUs select the MIMO SUs as cooperative relays; and what are the practical deployment scenarios for MIMO-CCRN. Specifically, the contributions of this section of dissertation are three-fold:

1) We propose a novel MIMO-CCRN system architecture. Considering the DoFs of the nodes, we schedule the transmissions in both spatial and temporal domains to improve spectral efficiency.

2) We consider the case of one primary link and multiple SUs. The beamforming design and the link capacities are studied. We then analyze the problem under a Stackelberg game framework. A unique Nash Equilibrium is achieved which provides the optimal strategy.

3) We consider the practical use case of MIMO-CCRN in a cellular network with multiple MIMO-empowered femtocells. To model network-wide cooperation and competition, we design an algorithm to find a stable matching of all the PUs and SUs in the network.
3.2 Preliminaries: MIMO Characteristics

In this section, we briefly explain the basics of MIMO and its benefits. Since MIMO is a broad category containing various techniques, we will mainly focus on introducing Zero-Forcing Beam-Forming (ZFBF), which is intensively used in our MIMO-CCRN framework design.

3.2.1 Zero-Forcing Beam-forming

ZFBF is one of the most powerful interference mitigation techniques in MIMO systems [85, 86]. It uses multiple antennas to steer beams towards the intended receivers to increase the Signal-to-Noise Ratio (SNR), while forming nulls at unintended receivers to avoid interference. Such beamforming can be performed on both transmitter and receiver sides through appropriate pre- and post-coding on the signals. Since ZFBF performs linear correlation/ decorrelation with low complexity, it provides a tractable solution suitable for use in many MIMO-based cross-layer designs [77, 78, 87].

For ease of explanation, let us start with the standard $2 \times 2$ MIMO channel to understand the rationale of ZFBF, as shown in Figure 3.3. Two streams, $s_1$ and $s_2$, can be transmitted simultaneously through this MIMO link without interference. Before transmission, precoding can be performed on the two streams by multiplying the stream $s_i$ with an encoding vector $u_i = [u_{i1} \ u_{i2}]^T$. Therefore, the resulting transmitted signal will be $s_t = u_1 s_1 + u_2 s_2$. Each antenna transmits a weighted combination of the original stream $s_1$ and $s_2$. 

![Figure 3.3: Transmission of two streams: a $2 \times 2$ MIMO channel.](image)
Let $H_{t,r}$ denote the $2 \times 2$ channel matrix between the transmitter and the receiver. Each entry $h_{ij}$ of $H_{t,r}$ is a complex channel coefficient along the path from the $j^{th}$ antenna on the transmitter to the $i^{th}$ antenna on the receiver. Therefore, we can represent the received signals on the receiver side as:

$$s_r = H_{t,r} s_t + n = H_{t,r} u_1 s_1 + H_{t,r} u_2 s_2 + n$$  \hspace{1cm} (3.1)$$

where $n$ is the i.i.d. $\mathcal{C}\mathcal{N}(0, \sigma^2 I_2)$ channel noise. Since the receiver has two antennas, representing the signals as 2-dimensional vectors is convenient \[84\]. We can see that the receiver actually receives the sum of two vectors which are along the directions of $H_{t,r} u_1$ and $H_{t,r} u_2$. The encoding vectors $u_1$ and $u_2$ control the direction of the vectors.

Eqn. 3.1 shows that the two streams interfere with each other on the receiver side. A method to remove such inter-stream interference is to project the received signal $s_r$ onto the subspace orthogonal to the one spanned by the other signal vector. Specifically, we can apply two decoding vectors $v_1$ and $v_2$ on $s_r$ to decode $s_1$ and $s_2$ respectively as

$$\tilde{s}_i = v_i \dagger H_{t,r} u_1 s_1 + v_i \dagger H_{t,r} u_2 s_2 + v_i \dagger n \quad i = 1, 2$$  \hspace{1cm} (3.2)$$

If we judiciously configure the encoding and decoding vectors in a way that $v_1 \dagger H_{t,r} u_2 = 0$ and $v_2 \dagger H_{t,r} u_1 = 0$, the two streams $s_1$ and $s_2$ can be decoded without interference. ZFBF can thus realize spatial multiplexing of the streams. In the situations where the co-channel interference is much stronger than the noise, the channel capacity can be significantly improved.

### 3.2.2 ZFBF in multi-user MIMO scenarios

The above example shows how to manipulate the encoding and decoding vectors to nullify the interference in a single user-pair case. More often than not, ZFBF is adopted as an interference mitigation technique in multi-user MIMO scenarios \[85\][86], like cellular uplink/downlink. We will briefly illustrate it in the context of cognitive radio networks as follows, which is also the model presented in \[83\].
The right part of Figure 3.4 shows an example in which ZFBF improves the spatial reuse of the channel with multiple users. Consider that a pair of PUs, each equipped with one antenna, forms a primary link. A pair of SUs forms a secondary link with each SU equipped with two antennas. The primary link and secondary link are within each other’s interference range. The channel coefficient matrices between different transmitter/receiver combinations are denoted as $h_{PT,PR}$, $h_{PT,SR}$, $h_{ST,PR}$ and $H_{ST,SR}$. Note that depending on the number of transmitting and receiving antennas, their dimensions are $1 \times 1$, $2 \times 1$, $1 \times 2$ and $2 \times 2$ respectively.

Two independent streams, one primary stream $s_p$ and one secondary stream $s_s$, can be transmitted simultaneously. Suppose the encoding and decoding vectors applied on the secondary link are $u_s$ and $v_s$, the final signals on both primary and secondary receivers are

$$
\tilde{s}_p = h_{PT,PR} s_p + h_{ST,PR} u_s s_s + n_p
$$

$$
\tilde{s}_s = v_s^\dagger h_{PT,SR} s_p + v_s^\dagger H_{ST,SR} u_s s_s + v_s^\dagger n_s
$$

(3.3)

If we intentionally configure $u_s$ and $v_s$ so that $h_{ST,PR} u_s = 0$ and $v_s^\dagger h_{PT,SR} = 0$, both primary and secondary signals can be decoded at their corresponding receivers. In this example, the co-channel interference is suppressed due to ZFBF. The spatial reuse factor is improved by letting two interfering links transmit simultaneously, where the PUs’ transmission is not affected as the SUs access the channel. The general case of the capacity of multi-user MIMO based on ZFBF is studied in [86]. Note the encoding/decoding vectors commonly have unit length.
3.2.3 Remarks on employing ZFBF

Although ZFBF can provide appealing benefits, several issues need to be carefully considered when employing it, which are discussed below:

1. To properly configure the encoding and decoding vectors, both transmitter and receiver should be aware of the instantaneous channel coefficient matrix. This is a common assumption in [33, 74, 86]. However even without such an assumption, there exist practical estimation techniques already being applied in implementations which give fairly good results [88].

2. The ability of ZFBF to enable spatial multiplexing and suppress interference, is not unlimited. Fundamentally, the number of concurrent streams that can be scheduled is constrained by the DoF of the transmitting node. Also, the number of streams a receiver can simultaneously receive is also limited by its DoF [79]. We will carefully consider the nodes’ DoFs in scheduling the transmissions in MIMO-CCRN.

3.3 The Case of One Primary Link and Multiple SUs

In this section, we study MIMO-CCRN in the case of one primary link and multiple SUs. Due to the practical constraint on the distances between antennas to ensure independent fading, we will mainly consider the case that the SUs are equipped with two antennas. The general case of SUs equipped with multiple antennas is also discussed.

3.3.1 System Model

We consider a secondary network consisting of \( K = |S| \) transmitter-receiver pairs, each of which is denoted by \((ST_i, SR_i), i \in S\). SUs are equipped with two MIMO antennas. Each PU is assumed to be a legacy device with a single antenna. They are co-located and all the entities interfere with each other. The primary transmission is divided into frames and we use \( T \) to represent the frame duration (FD). The primary link can select
a subset of secondary pairs to participate in the cooperative transmission, denoted as \( \mathcal{R} \). Either ST or SR in each pair can be the relay. Note that this gives more flexibility to PUs’ relay selection as compared to \([33, 73]\), in which only STs can be chosen as the relay. Figure 3.5 demonstrates the frame structure we use. If cooperative communication is enabled, a FD is time-divided into two phases. In the first phase with duration \( \alpha T \), the Primary Transmitter (PT) broadcasts the primary data to the secondary relays in \( \mathcal{R} \). Then in the second phase with duration \( (1 - \alpha)T \), those secondary relays cooperatively transmit the data to the Primary Receiver (PR). We define \( \alpha = 1 \) as a special case when PT uses the entire FD for a direct transmission to PR without cooperation. We can see that compared to the existing CCRN schemes, MIMO-CCRN totally avoids a time fraction dedicated for the SUs’ transmissions. In return for the SUs’ cooperation, the channel will be granted to the relays for their own traffic. As a result of the use of MIMO, their transmissions can be scheduled into two phases. The detailed procedure is illustrated next.

**Phase One**

Figure 3.5 shows the system architecture. In this example, there are \(|S| = 4\) pairs of SUs. We suppose the primary link selects SR\(_1\), SR\(_2\), ST\(_3\) and ST\(_4\) as the cooperative relays, which are marked in black in the figure. Throughout Phase One, PT continuously broadcasts its data to the chosen relays. For the secondary network, the pairs with SR selected as the relay are allowed to access the channel in this phase in a TDMA fashion. In our example, they are the pairs (ST\(_1\), SR\(_1\)) and (ST\(_2\), SR\(_2\)). We use \( S_1 \) to denote the set of such pairs. Thus Phase One is further divided into \(|S_1| \) subslots (\(|S_1| = 2\) in our example), one for each pair. In a symmetric way, the pairs with ST selected as the relay, denoted by \( S_2 \), are granted access the channel in Phase Two. It is obvious that \( S_1 \) and \( S_2 \) are disjoint sets and \( S_1 \cup S_2 = \mathcal{R} \subseteq S \).

We use \( h_{0r} \) to represent the channel coefficient vector from PT to the relay node \( r \), \( \forall r \in \mathcal{R} \). Note this node stands for SR\(_r\) if \( r \in S_1 \) and ST\(_r\) if \( r \in S_2 \). Also \( H_{ir} \) is used to represent the channel coefficient matrix from ST\(_i\) to the relay node \( r \), \( \forall i \in S_1, r \in \mathcal{R} \). Suppose a subslot of length \( T_k^{(1)} \) is allocated to the pair (ST\(_k\), SR\(_k\)), \( k \in S_1 \) in Phase
One. By virtue of multiple antennas, SR\(_k\) can receive both streams from PT and ST\(_k\) simultaneously in this subslot. We further denote the primary stream as \(s_p\) and the stream transmitted by ST\(_k\) in this subslot as \(s_k\). If ST\(_k\) applies an encoding vector \(u_k^{(s)}\) on \(s_k\), then the received signal on each relay \(r\) is the combination of PT’s stream and ST\(_k\)’s stream,

\[
s_{r,k}^{(rec)} = h_{0r}s_p + H_{kr}u_k^{(s)}s_k + n, \quad \forall r \in \mathcal{R}, k \in \mathcal{S}_1,
\]

which can be viewed as the combination of two vectors in a two-dimensional space. Then each relay \(r\) can apply a decoding vector \(v_r^{(p)}\) to decode the primary stream, by letting \(v_r^{(p)\dagger}H_{kr}u_k^{(s)} = 0\). The resulting primary signal on \(r\) is then

\[
\tilde{s}_{r,k}^{(p)} = v_r^{(p)\dagger}h_{0r}s_p + v_r^{(p)\dagger}n, \quad \forall r \in \mathcal{R}, k \in \mathcal{S}_1.
\]

Being one of the relays, SR\(_k\) uses another decoding vector \(v_k^{(s)}\) to decode the secondary stream for itself. By letting \(v_k^{(s)\dagger}h_{0k} = 0\), its own stream sent by ST\(_k\) can be decoded as

\[
\tilde{s}_k = v_k^{(s)\dagger}H_{kk}u_k^{(s)}s_k + v_k^{(s)\dagger}n, \quad k \in \mathcal{S}_1.
\]

Therefore we can clearly see that in Phase One, the relays continuously receive the primary data from the PT, meanwhile those pairs in set \(\mathcal{S}_1\) perform their own transmissions in their respective subslots.
Phase Two

In Phase Two, a similar idea can be applied as in Phase One. The selected relays cooperatively forward the primary data to the PR, meanwhile, the pairs in the set $S_e$ will access the channel in a TDMA fashion. As in Figure 3.5, $(ST_3, SR_3)$ and $(ST_4, SR_4)$ share the channel by dividing it into two subslots, one for each pair.

We use $H_{ri}$ to denote the channel coefficient matrix from relay $r$ to $SR_i, \forall r \in \mathcal{R}, i \in S_e$, and $h_{r0}$ is used to represent the channel coefficient vector from relay $r$ to PR. Also the node $r$ stands for $SR_r$ if $r \in S_\infty$ and $ST_r$ if $r \in S_e$. Without ambiguity, we still use $s_p$ and $s_k$ to denote the primary stream and the secondary stream $ST_k$ sends. Suppose a subslot of length $T_k^{(2)}$ is allocated to the pair $(ST_k, SR_k), k \in S_e$. Since ST$_k$ has multiple antennas, it can transmit both primary and secondary streams to the PT and SR$_k$ respectively without interference. Specifically, each relay $r$ (including ST$_k$) transmits $s_p$ encoded by $u_r^{(p)}$. Meanwhile, ST$_k$ also transmits its own signal $s_k$ encoded with vector $u_k^{(s)}$, which is combined with the primary signal it sends. If $u_k^{(s)}$ is chosen so that $h_{k0}u_k^{(s)} = 0$, the secondary stream from ST$_k$ is totally nulled at the PR. The signal received by PR is then

$$\tilde{s}_p = \sum_{r \in \mathcal{R}} \sqrt{P_r} h_{r0} u_r^{(p)} s_p + n.$$ \hspace{1cm} (3.7)

Moreover, the received signal for SR$_k$ in this subslot is

$$s_k^{(rec)} = \sum_{r \in \mathcal{R}} \sqrt{P_r} H_{rk} u_r^{(p)} s_p + H_{kk} u_k^{(s)} s_k + n,$$ \hspace{1cm} (3.8)

$\forall r \in \mathcal{R}, k \in S_e$. We use $P_r, r \in \mathcal{R}$ to denote the transmission powers used for relaying by relay $r$. The exact values of $P_r$’s are determined by the secondary power control game described in Section 3.3.5. The first term in Eqn. (4.8) is the primary signal summed over all the relays. The second term is the secondary signal transmitted by ST$_k$. The secondary signal $s_k$ can thus be easily decoded by choosing a decoding vector $v_k^{(s)}$ such that the primary signal can be canceled. This results in

$$\tilde{s}_k = v_k^{(s)\dagger} H_{kk} u_k^{(s)} s_k + v_k^{(s)\dagger} n, \quad k \in S_2.$$ \hspace{1cm} (3.9)

In summary, in Phase Two the relays continuously forward the primary data to the PR, meanwhile those pairs in $S_2$ perform their own transmissions in their respective subslots.
We will study how to determine the length of each subslot in Phase One and Two, $T^{(1)}_k$ and $T^{(2)}_k$, in Section 3.3.5.

### 3.3.2 Link Data Rate Analysis

Based on the previous model, the data rates for both primary and secondary links can be resolved.

**Primary Link**

For the cooperative communication, we assume the use of a collaborative scheme based on decode-and-forward (DF) due to its simplicity in presentation, and at the receiving end, the PR exploits maximum ratio combining (MRC) before decoding the signal. Our scheme can be extended to use other cooperation techniques to obtain a greater achievable primary rate.

In Phase One, since there are multiple relays in the downlink, the rate is easily shown to be dominated by the worst channel in the subset $r \in \mathcal{R}$. Suppose the transmission power of PT is $P_P$, then according to Eqn. (3.5), in the subslot when ST$_k$ is transmitting, the downlink rate is

$$R^{(PS)}_k = \log_2(1 + \min_{r \in \mathcal{R}} |v_{r,k}^{(p)} h_{0r}^0|^2 P_P N_0), \ k \in S_1.$$  

(3.10)

In Phase Two, since MRC is used, the effective SNR at PR equals to the sum of the SNRs from all the secondary relays. Based on Eqn. (4.7), in the subslot when ST$_k$ is transmitting, the achievable rate of the cooperative link is given by

$$R^{(SP)}_k = \log_2(1 + \sum_{r \in \mathcal{R}} \frac{|h_{r0}^{(p)} u_{r,k}^{(p)}|^2 P_r}{N_0}), \ k \in S_2.$$  

(3.11)

Moreover, denote the channel gain from PT to PR as $h_P$. Then in the trivial case when the secondary cooperation is not applied, the rate of the direct transmission from PT to PR is

$$R_{dir} = \log_2(1 + \frac{|h_P|^2 P_P}{N_0}).$$  

(3.12)
Secondary Link

For simplicity, it is assumed that for each secondary pair, ST will adopt a fixed power level for transmitting the secondary data, while the power $P_r$ used for relaying the primary data is adaptive. Denote $P_k^{(s)}$ to be the power used by ST$_k$ for secondary data transmission. Based on Eqn. (3.6) and Eqn. (3.9), the transmission rate of secondary link (ST$_k$, SR$_k$) in the two phases can be unified as

$$R_k^{(s)} = \log_2(1 + |v_k^{(s) \dagger} H_{kk} u_k^{(s)}|^2 P_k^{(s)}) \quad \forall k \in \mathcal{R}. \quad (3.13)$$

Specifically, in Phase One, for each ST$_k$, $v_k^{(s)}$ can be chosen to satisfy $v_k^{(s) \dagger} \mathbf{h}_0 = 0$. Then $u_k^{(s)}$ can be chosen in the same direction as $v_k^{(s) \dagger} H_{kk} u_k^{(s)}$ to maximize $|v_k^{(s) \dagger} H_{kk} u_k^{(s)}|^2$. Accordingly, $v_{r,k}^{(p) \dagger}$ can be resolved for each relay $r$ given $u_k^{(s)}$. In Phase Two, $u_k^{(s)}$ is chosen to let $\mathbf{h}_{k0} u_k^{(s)} = 0$. Since the $P_r$’s cannot be determined a priori in Eqn. (4.7), we will align each $H_{rk} u_{r,k}^{(p)}$ in the same direction that is orthogonal to $H_{kk} u_k^{(s)}$. As a result, $u_k^{(p)}$ can be resolved to satisfy $(H_{kk} u_k^{(s)})^\dagger H_{rk} u_{r,k}^{(p)} = 0$. Also, given $u_k^{(s)}$, $v_k^{(s)}$ is computed to maximize $|v_k^{(s) \dagger} H_{kk} u_k^{(s)}|^2$.

To conclude, given the sets of relays $S_1$, $S_2$ and the channel matrices, all the encoding/decoding vectors can be determined, thus the primary link rates $R_k^{(PS)}$ and $R_k^{(SP)}$ are resolved. Further, all the relay pairs can locally calculate the rate $R_k^{(s)}$ for its own transmission.

3.3.3 Problem Formulation

We consider the objective of the primary link is to maximize its utility, termed as throughput, over the different combinations of relay sets $S_1$, $S_2$, and the time length scale $\alpha$ of the two phases. The throughput for cooperative communication is the minimum of the throughput in the two phases:

$$R_{coop} = \min\left\{ \sum_{i \in S_1} T_i^{(1)} R_i^{(PS)}, \sum_{i \in S_2} T_i^{(2)} R_i^{(SP)} \right\}. \quad (3.14)$$
So the primary rate $R_P$ in this frame duration is simply

$$R_P = \begin{cases} R_{\text{dir}} & \alpha = 1 \\ R_{\text{coop}} & 0 < \alpha < 1 \end{cases}. \quad (3.15)$$

Thus the primary link aims at solving the following primary utility maximization problem:

$$\max_{\alpha, S_1, S_2, T_i^{(1)}, T_i^{(2)}, P_r} R_P, \quad \text{Subject to:} \quad \begin{align*}
\sum_{i \in S_1} T_i^{(1)} &= \alpha T, \quad 0 < \alpha \leq 1. \\
\sum_{i \in S_2} T_i^{(2)} &= (1 - \alpha)T, \\
0 &\leq P_r \leq P_r^{\text{max}}, \forall r \in \mathcal{R}, \\
S_1, S_2 &\subseteq \mathcal{S} \text{ and } S_1 \cap S_2 = \emptyset. \quad (3.16)
\end{align*}$$

The first and second constraint limits the total length of the subslots in Phase One and Phase Two. The third constraint means the transmission power for relaying the primary signal of each relay $r$ is bounded by $P_r^{\text{max}}$, the power budget for relaying. Due to the selfish nature of the secondary network, $P_r$’s are determined as the result of the competition between the SUs. This will be illustrated in detail in Section V. The set $\mathcal{R} = S_1 \cup S_2$ is determined once the sets $S_1$ and $S_2$ are known.

### 3.3.4 Beyond Two Antennas

In the general case of multiple antennas per SU in MIMO-CCRN, the principle for relaying the primary data remains the same. Besides, multiple concurrent data streams can be transmitted between a secondary pair by using spatial multiplexing. For example, when all the SUs are equipped with three antennas in Figure 3.5, $ST_1$ can simultaneously transmit two streams for its own traffic to $SR_1$ in its subslot, the same is true for $ST_2$, $ST_3$ and $ST_4$ in their respective subslots. Generally, to make the streams decodable, the number of concurrent streams a node can transmit should be no more than its DoF, which is given by the number of antennas it has. Symmetrically, the number of streams a receiver can simultaneously receive (including the interfering streams) is also limited by its DoF \[79\]. This fact characterizes the feature that MIMO increases the link capacity linearly with the number of antennas.
Guided by the above principle, we discuss the feasibility of link-layer stream scheduling for the secondary network. The details of computing the beamforming vectors are omitted here. We define the number of antennas of ST$_i$ and SR$_i$ as $Ant_{ST_i}$ and $Ant_{SR_i}$ respectively. We assume PT and PR have one antenna each and a relay set $\mathcal{R}$ is given. In Phase One, when ST$_k$ is scheduled to transmit in its subslot, it should guarantee the number of streams other relays receive does not exceed their DoFs. Therefore, the number of secondary streams it can send is $str_k = \min_{\forall r \in \mathcal{R}}\{Ant_{ST_k}, Ant_r - 1\}$. The decrease by one of $Ant_r$ is due to the reception of the primary stream. Similarly in Phase Two, in the subslot for ST$_k$ to transmit, ST$_k$ should relay one primary stream, while SR$_k$ is receiving an interfering primary stream. Thus $str_k = \min\{Ant_{ST_k} - 1, Ant_{SR_k} - 1\}$ streams can be sent by ST$_k$ for its own traffic.

### 3.3.5 Game Theory Analysis

In this section, we analyze our problem under a typical two-stage Stackelberg game framework. We will resolve the unique Nash Equilibrium (NE) for the secondary power control game, and maximize the primary link’s utility based on the knowledge of NE.

**Secondary Power Control Game**

In the context of spectrum leasing in CCRN, both PUs and SUs are intrinsically selfish. It is therefore best to analyze the problem using Stackelberg game [33][73]. The PU owns the spectrum and thus is the leader with a higher priority in choosing the optimal relay sets and parameters. The secondary pairs in $\mathcal{S}$ are the followers competing with each other to decide the best strategy to share the spectrum. All the entities are rational, aiming to maximize their own utilities. Guided by the idea of backward induction [33][73], it is necessary to decompose the problem so that the optimal $T_i^{(1)}$, $T_i^{(2)}$ and $P_r$ in (3.16) can be obtained if $S_{\infty}$, $S_{\epsilon}$ and $\alpha$ are given. This is achieved by finding a unique NE for the secondary power control game. Then based on the knowledge of the NE, the primary links determines the best relay sets $S_{\infty}$, $S_{\epsilon}$ and the parameter $\alpha$. 
In the current model, secondary pairs compete with each other for the channel access, in terms of the durations of the subslots in Phase One and Phase Two. For each secondary pair \( k \in \mathcal{R} \), the utility function is defined as the difference between the achievable throughput and the cost of energy used in this frame duration as in [33], which is then:

\[
 u^{(s)}_k = T^{(i)}_k (R^{(s)}_k - wP^{(s)}_k) - wP_k (1 - \alpha) T, \ \forall k \in S_i, \tag{3.17}
\]

where \( R^{(s)}_k \) is determined by Eqn. (3.13), \( w \) is the cost per unit transmission energy and \( P_k \) is the power used for relaying adopted by the secondary pair \( k \).

Meanwhile, we let \( T^{(1)}_k \) and \( T^{(2)}_k \) be proportional to relay \( k \)'s consumed energy for relaying, for fairness, which is represented as

\[
 T^{(i)}_k = c_i \cdot \frac{P_k}{\sum_{j \in S_i} P_j}, \tag{3.18}
\]

where \( c_i = \alpha T \) for \( k \in S_1 \) and \( c_i = (1 - \alpha) T \) for \( k \in S_2 \). We can see that the utility function for each secondary pair is a function of their transmission power used for the primary signal relaying, therefore a secondary power control game can be formulated. Secondary pairs in each set \( S_i \) are the players, and form a non-cooperative power selection game where they compete in the same set to maximize their own utility. The strategy space is the power \( \mathcal{P} = [P_k] : 0 \leq P_k \leq P^\text{max}_k \). The best strategy can be resolved for each relay when the NE is achieved. Based on Eqn. (3.17) and (3.18) and using \( \hat{R}^{(s)}_k \) to replace \( R^{(s)}_k - wP^{(s)}_k \), the utility for the secondary pair \( k \) in \( S_1 \) is

\[
 u^{(s)}_k = \alpha T \cdot \frac{P_k}{\sum_{i \in S_\infty} \hat{R}^{(s)}_i} - wP_k (1 - \alpha) T, \ k \in S_1. \tag{3.19}
\]

In this section, we analyze in detail the NE for the secondary pairs in \( S_1 \) based on Eqn. (3.19). Similar methods can be applied to the game among the relays in set \( S_2 \). We will first prove the existence and uniqueness of the Nash Equilibrium.

**Theorem 1**: A Nash Equilibrium exists in the secondary power control game.

**Proof**: Note that Eqn. (3.19) has similar form as the utility function defined in [73] (Eqn. (7)). Using the same method, we can first prove that \( P_k \) is a nonempty, convex and compact subset of the Euclidean space \( \mathbb{R} \), then prove that \( u^{(s)}_k \) is continuous and concave.
in $P_k$. A Nash Equilibrium then exists if these two conditions satisfy. We omit the details, and let interested readers refer to [73].

To analyze the uniqueness of the equilibrium, we should refer to the best response function of player $k$ given the power selection of other players. Since the utility function $u_k^{(s)}$ is concave, the best response is achieved when the first derivative of $u_k^{(s)}$ with $P_k$ equals to 0, as

$$\frac{\partial u_k^{(s)}}{\partial P_k} = \frac{\alpha T \hat{R}_k^{(s)} \sum_{i \in S_1, i \neq k} P_i}{(\sum_{i \in S_1} P_i)^2} - w(1 - \alpha)T = 0. \quad (3.20)$$

Solve Eqn. (3.20) and eliminate the trivial cases when the power is negative or exceeds $P_k^{max}$, the best response function is

$$r_k(P) = \sqrt{\frac{\alpha \hat{R}_k^{(s)} \sum_{i \in S_1, i \neq k} P_i}{w(1 - \alpha)} - \sum_{i \in S_1, i \neq k} P_i}, \quad (3.21)$$

$\forall k \in S_1$, with the following constraint:

$$0 \leq r_k(P) \leq P_k^{max}. \quad (3.22)$$

**Theorem 2**: The secondary power control game has a unique Nash Equilibrium.

**Proof**: It amounts to proving that the system represented by the equation set (3.21) has a unique solution. Solving the equation set (3.21) consisting of $|S_\infty|$ equations, the resulting relaying power for SR$_k$ when $k \in S_1$ is

$$P_k^* = \frac{\alpha}{1 - \alpha} A_k, \quad (3.23)$$

where

$$A_k = \frac{(|S_1| - 1) \hat{R}_k^{(s)} \sum_{i \in S_1 \hat{R}_i^{(s)}}}{w \sum_{i \in S_1} \hat{R}_i^{(s)}} (1 - \frac{|S_1| - 1}{\sum_{i \in S_1} \hat{R}_i^{(s)}}).$$

We can see that the resulting $P_k^*$ is unique for each relay in $S_1$ which is the transmitting power it will adopt to relay the primary data when the equilibrium is reached.
Similarly, we can prove that the NE point also exists and is unique for secondary power control game among the secondary relay pairs in the set $S_2$. The relaying power for each pair should be chosen as

$$P^*_k = B_k = \frac{(|S_2| - 1)}{w \sum_{i \in S_2} \frac{1}{R^{(s)}_k}} (1 - \frac{|S_2| - 1}{R^{(s)}_k \sum_{i \in S_2} \frac{1}{R^{(s)}_i}}).$$  (3.24)

Note that $P^*_k$ is independent of $\alpha$ for relay pairs which belong to $S_2$. In Section V, a unique NE point is found for all the simulations for each set $S_1$ and $S_2$.

**Maximizing the Primary Link’s Utility**

Based on the analytical result of the secondary power control game, as the leader of the Stackelberg game, the primary link can resolve the best system parameters to solve the formulated primary utility maximization problem.

The relaying power for each relay can be obtained according to Eqns. (3.23) and (3.24), where $a_k$ and $b_k$ are known if $S_1$, $S_2$ and the secondary link rates are given. To resolve the optimal $\alpha$, we can substitute (3.23) and (3.24) into (3.11), and the resulting link rate in Phase Two of MIMO-CCRN is

$$R^{(SP)}_i = \log_2(1 + A_i \cdot \frac{\alpha}{1 - \alpha} + B_i), \ i \in S_\infty,$$  (3.25)

where $A_i = \sum_{k \in S_1} \frac{|h_{iou}^{(p)}|^2 a_k}{N_0}$, $B_i = \sum_{k \in S_2} \frac{|h_{iou}^{(p)}|^2 b_k}{N_0}$ and $r$ refers to SR$_k$ when $k \in S_1$, and ST$_k$ when $k \in S_2$.

Moreover, it has been proved that in two-phase cooperative communications [33], the throughput is maximized when the downlink throughput equals the uplink throughput. In our problem, to maximize the PU’s throughput, we should have:

$$\sum_{k \in S_1} T^{(1)}_k R^{(PS)}_k = \sum_{k \in S_2} T^{(2)}_k R^{(SP)}_k.$$  (3.26)

Expanding Eqn. (3.26) based on (3.18) and (3.23), we have:

$$\sum_{k \in S_2} D_k \log_2(1 + A_k \cdot \frac{\alpha^*}{1 - \alpha^*} + B_k) = C \cdot \frac{\alpha^*}{1 - \alpha^*}.$$  (3.27)
where $C = \sum_{k \in S_1} \frac{P_k^* R_k^{(ps)}}{P_k^*} = \sum_{k \in S_1} \frac{a_k R_k^{(ps)}}{a_k}$ and $D_k = \frac{P_k^*}{P_i^*} = \frac{b_k}{b_i}$. $\alpha^*$ in the above formula is the optimal $\alpha$. It is easy to see that $\frac{\alpha^*}{1-\alpha^*}$ is the x-coordinate of the intersection point between the summation of a set of log functions and a straight line passing through the origin. Thus, the value of $\alpha^*$ can be efficiently found using binary search.

Based on the the above, and given $S_1$ and $S_2$, $\alpha^*$ can be resolved, which determines the optimal durations of the two phases to maximize the throughput of the primary link. In a practical implementation, the secondary network measures the channel coefficient matrices $h_{0r}, h_{r0}, H_{ir}$ and $H_{ri}$, while PR measures $h_p$. Then PT periodically collects this data. From the universal set of the relay pairs $S$, the PT can enumerate all the possible sets $S_1$ and $S_2$ which satisfy the criteria (3.22). From all the possible sets, the one that maximizes the primary link’s utility can be selected. The information consisting of the optimal set $R$, $\alpha^*$ and beam-forming vectors will be piggybacked to the SUs. The secondary pairs being selected as the relays can calculate the best relaying power in a distributed fashion.

### 3.4 The Case of MIMO-empowered Femtocells

Next, we study a scenario consisting of a cellular network and multiple MIMO-empowered femtocells. We aim to show the potential of MIMO-CCRN in this typical practical scenario. However, we envisage there also exist more use cases for MIMO-CCRN.

As shown in Figure 3.6, there exist a group of femtocell networks under the umbrella of a macrocell base station (BS). There are also mobile user equipments (UEs) in the network. Typically, UEs are allocated a set of licensed channels for data access via frequency-division. Thus we define the link between the macrocell BS and a UE as the primary link, such as the link BS-UE1 and BS-UE2 in the figure. Each femtocell access point is assumed to have multiple antennas and to serve its own local network. It may have the cognition ability and can sense and adapt to different channels, as stated in [89]. We call it a secondary access point (Femto-SAP, or SAP) and its network a secondary network (SN). For the purpose of interference management, normally the SAP is only allocated a small portion of the licensed spectrum [90], which limits its performance. Under MIMO-CCRN,
it can dynamically relay the traffic for the UEs, while borrowing their spectrum.

### 3.4.1 System Model

In a normal cellular environment, there exist multiple UEs and SAPs. As here we mainly study from a network point of view, we simplify the relay model by assuming one UE recruits at most one Femto-SAP as the relay. We also assume a frame duration is equally divided into the two phases. Similar to Section 3.3, in Phase One, the SAP can select an SU, denoted as SU\textsubscript{u}, to receive its uplink secondary traffic, while receiving the primary stream. In Phase Two, the SAP forwards the data to the UE, and it also selects an SU, SU\textsubscript{d}, to transmit the downlink secondary traffic to. This is illustrated as in Figure 3.7. We still consider the case where the SAPs are equipped with two antennas.

Let us first focus on Phase One. Without losing generality, we assume that the licensed spectrum band owned by UE consists of a set of orthogonal subchannels, denoted by $\mathcal{C}_{UE}$. Once a UE is partnered with a SAP, it leases all its subchannels to the SAP in exchange for cooperation. For a subchannel $c \in \mathcal{C}_{UE}$, we use $h_{ts}$ and $h_{us}^c$ to represent the channel gains of the link BS-SAP and SU\textsubscript{u}-SAP on $c$. we denote the primary stream sent from BS as $s_p$ and the stream transmitted from SU\textsubscript{u} as $s_u$. The received signal on the SAP on channel $c$ is
then:

\[ s^{(r)}_{SAP} = h_{ts}^c s_p + h_{us}^c s_u + n, \] (3.28)

This is a simpler case of the model we have analyzed in Sec. III. Using ZFBF, the channel gain on channel \( c \) for the primary stream will be equivalent to \( \tilde{h}_{ts}^c = v_u^c h_{ts}^c \), where \( v_u^c \) is the decoding vector. Due to the space limit, we omit the derivation details and let interested readers refer to [70]. Further we use \( P_{ts}^c \) to represent the transmission power that the BS allocates on subchannel \( c \). The maximal achievable transmission rate \( R_{ts} \) from the BS to the SAP can be found by solving:

\[
\max_{P_{ts}^c \geq 0} R_{ts} = \sum_{c \in C_{UE}} \log(1 + \frac{P_{ts}^c |\tilde{h}_{ts}^c|^2}{N_0}),
\] subject to:

\[
\sum_{c \in C_{UE}} P_{ts}^c \leq P_{ts}^{max},
\]  

(3.29)

where \( P_{ts}^{max} \) is the maximal transmission power allowed by the BS for this UE. The solution to problem (3.29) can be readily achieved by the waterfilling strategy [84]. The optimal power allocation is given by:

\[
P_{ts}^{cs} = \left( \frac{1}{\lambda} - \frac{N_0}{|\tilde{h}_{ts}^c|^2} \right)^+,
\]  

(3.30)

where \( x^+ = \max(x, 0) \). The parameter \( \lambda \) is chosen to meet \( \sum_{c \in C_{PT}} P_{ts}^{cs} = P_{ts}^{max} \), which can be resolved efficiently using binary search. In the following, we denote this waterfilling solution as \( R_{ts} = WF(P_{ts}^{max}, \tilde{h}_{ts}^c, C_{UE}) \). \( \tilde{h}_{ts}^c \) is the equivalent channel gain vector for the subchannels of \( C_{UE} \).
Further we use $R_{us}$, $R_{sr}$, $R_{sd}$ to denote the maximal achievable rates on the links SU$_u$-SAP, SAP-UE and SAP-SU$_d$ respectively by exploiting cooperation. Also, $\tilde{h}_{us}^c = v_u^c h_{us}^c$, $\tilde{h}_{sr}^c = h_{sr}^c u_p^c$ and $\tilde{h}_{sd}^c = h_{sd}^c u_d^c$ are used to represent the equivalent channel gains for subchannels $c$ on each of these links. To be more general, we consider that the SAP already has its own set of available subchannels, denoted by $C_{SAP}$. For example, a basic set of channels pre-allocated by its service provider. The interference on these subchannels is assumed to be minimal or known in advance so no beamforming is necessary on these subchannels. Therefore, similarly we can derive:

$$R_{us} = \mathcal{WF}(P_{max}^u, \tilde{h}_{us}^c, C_{BS} \cup C_{SAP}),$$ (3.31)

$$R_{sr} = \mathcal{WF}(P_{sr}, \tilde{h}_{sr}^c, C_{BS}),$$ (3.32)

$$R_{sd} = \mathcal{WF}(P_{sd}, \tilde{h}_{sd}^c, C_{BS} \cup C_{SAP}).$$ (3.33)

In the above, $P_{max}^u$ is the maximal transmission power of SU$_u$, and $P_{sr}$ and $P_{sd}$ are the transmission powers that the SAP allocates for the primary and secondary streams. We require SAP to deliver all received data to its cooperating UE, then the following power and flow conservation constraints apply:

$$R_{sr} \geq R_{ts}, \quad P_{sr} + P_{sd} \leq P_{max}^s.$$ (3.34)

Then the SAP should first adjust $P_{sr}$ to meet the flow conservation constraint, and use its remaining power $P_{max}^s - P_{sr}$ to deliver its own traffic to SU$_d$.

Moreover, if we use $h_{tr} = \{h_{tr}^c, c \in C_{UE}\}$ to represent the channel gain vector for all the subchannels $c \in C_{UE}$ on the link BS-UE, the direct transmission rate from BS to the UE without cooperation is:

$$R_{tr} = \mathcal{WF}(P_{max}^p, h_{tr}^c, C_{UE}).$$ (3.35)

Similarly, the direct transmission rates from SU$_u$ to the SAP and SAP to the SU$_d$ is:

$$R_{us}^{dir} = \mathcal{WF}(P_{max}^u, h_{us}^c, C_{SAP});$$ (3.36)

$$R_{sd}^{dir} = \mathcal{WF}(P_{max}^s, h_{sd}^c, C_{SAP}).$$ (3.37)
where \( h_{us} = \{ \sqrt{|h_{us1}|^2 + |h_{us2}|^2}, c \in C_{SAP} \} \) and \( h_{sd} = \{ \sqrt{|h_{sd1}|^2 + |h_{sd2}|^2}, c \in C_{SAP} \}. \)

\( h_{usi}, h_{sdi}, i \in \{1, 2\} \) are the channel gain vectors from SU\(_u\) to the two antennas of the SAP and from the two antennas of the SAP to SU\(_d\). Maximal ratio combining is used to exploit antenna diversity.

### 3.4.2 Coalition Formation

Next we will focus on the cooperation among multiple UEs and MIMO-empowered SAPs, and aim to optimize the utilities of all the entities with a fairness guarantee.

We consider there exist \( N_P = |\mathcal{P}| \) UEs, each denoted as UE\(_i\), \( i \in \mathcal{P} \), and \( N_S = |\mathcal{S}| \) SNs, each of which is led by SAP\(_j\), \( j \in \mathcal{S} \). We define the utility each party can earn as a monotonically increasing function of its achievable data rate, i.e. \( U = U(R) \). We say a primary link \( i \) and secondary network \( j \) form a coalition when \( i \) and \( j \) cooperate. The resulting utility of UE\(_i\) is \( U_i^{p} = U_p(R_{ts}^{ij}) \), and the utility of the secondary network is \( U_j^{s} = U_s(w_uR_{us}^{ij} + w_dR_{sd}^{ij}) \). \( w_u \) and \( w_d \) are the weights SAP puts on its uplink and downlink. \( R_{ts}^{ij}, R_{us}^{ij} \) and \( R_{sd}^{ij} \) can be derived through Eqn. (3.32) to (3.33).

In practice, the secondary network measures the channel coefficients \( h_{ts}^c, h_{us}^c \) and \( h_{sd}^c, \forall c \in C_{UE} \cup C_{SAP} \). PRs measure the channel coefficients \( h_{ts}^c \) and \( h_{sr}^c \) for \( c \in C_{UE} \). We assume a common control channel is available for exchanging the messages. Then UEs and SAPs periodically exchange this data within their transmission range (local neighborhood). All the link rates can then be derived.

We consider the scenario without global information. Therefore, the coalitions can be only formed through local information exchange among the PLs and SNs. Under this setting, we find our problem is best modeled using two-sided matching theory [91, 92]. Based on two-sided matching theory, we define the following concepts in our coalition formation between the PLs and the SNs:

**Definition 1:** The entity \( k, k \in \mathcal{S} \cup \mathcal{P} \) is individually rational, if it will only form a coalition with others if such a coalition improves its utility, i.e. \( U_k,i > U_k^{dir}, \forall i \).

**Definition 2:** A blocking pair is a pair \((BS_i, UE_i)\) who both already have their respec-
tive partners \( n(i) \) and \( n(j) \), but prefer each other rather than their partners, i.e. \( U_{i,n(i)} < U_{i,j} \) and \( U_{j,n(j)} < U_{j,i} \).

We can easily see that if there exists a blocking pair, the entities involved have an incentive to break up from their current partnership and form a new coalition. Therefore, the current matching is \textit{unstable} and not desirable. The definition of matching \textit{stability} is given as:

\textit{Definition 3:} A matching is \textit{stable} if and only if every participating UE and SAP is individual rational, and there is no blocking pair in the network.

Based on above definitions, our objective is to find a stable matching in the primary and secondary market. Fortunately, it has been proved that a stable matching always exists for every two-sided market [91]. A \textit{distributed coalition formation algorithm} finds such a stable matching and is presented in the following algorithm. Our mechanism is an extension of [93].

In Algorithm 1, lines 1 \( \sim \) 12 shows the algorithm used by a UE \( p, p \in \mathcal{P} \). In the initialization step, \( p \) initializes a reference list by ranking the utilities it can obtain by cooperating with SAP \( j, U_{p,j}, j \in \mathcal{S} \) from the highest to the lowest. The SAPs that yield a utility lower than \( p \)’s direct transmission utility \( U_{p}^{\text{dir}} \) are ruled out from the list. As the algorithm starts, the initial status of \( p \) is set to \textit{free}. Whenever \( p \) is \textit{free} and has a non-empty preference list, it sends a “propose” message to the first SAP in its preference list, requesting for cooperation (lines 5 \( \sim \) 7). Then \( p \) waits for a message. If it receives an “accept” message from \( s \), it will set its partner to be \( s \). Otherwise, if it receives a “reject” message from any sender, \( p \) will remove the sender from its preference list (lines 10 \( \sim \) 11). If the sender of the “reject” message is its current partner, \( p \) will set its own status back to \textit{free} (lines 12 \( \sim \) 13).

Lines 14 \( \sim \) 16 show the algorithm used by an SAP \( s, s \in \mathcal{S} \). The setup of the preference list is similar to the UE side algorithm. \( s \) listens on the common control channel and waits for the messages from any UE. If it receives a “propose” message from \( p \), it will first check whether \( p \) is in its preference list. If not, \( s \) will reply with a “reject” message to \( p \) (line 19 \( \sim \) 20). If yes, \( s \) will set its partner to \( p \), and send \( p \) an “accept” message (lines
Algorithm 1: Coalition Formation Algorithm

1. foreach Primary Link \( p \in \mathcal{P} \) do
2. \( p \)’s preference list \( p\.list() \);
3. \( p\.partner \leftarrow \text{free}; \) end \( \leftarrow \text{false}; \)
4. while \( \text{end} \neq \text{false} \) do
5. if \( p\.partner = \text{free} \) and \( p\.list() \neq \emptyset \) then
6. \( s\leftarrow \text{pop}(p\.list()); \)
7. send a “propose” message to \( s; \)
8. if \( p \) receives a “accept” message from \( s \) then
9. \( p\.partner \leftarrow s; \)
10. if \( p \) receives a “reject” message then
11. delete the message sender from \( p\.list(); \)
12. if \( s \) is the message sender then
13. \( p\.partner \leftarrow \text{free}; \)

14. foreach SAP \( s \in \mathcal{S} \) do
15. \( s \)’s preference list \( s\.list() \);
16. \( s\.partner \leftarrow \text{free}; \) end \( \leftarrow \text{false}; \)
17. while \( \text{end} \neq \text{false} \) do
18. if \( s \) receives a “propose” message from \( p \) then
19. if \( p \notin s\.list() \) then
20. send a “reject” message to \( p; \)
21. else
22. send a “accept” message to \( p; \)
23. \( s\.partner \leftarrow p; \)
24. foreach \( t \) after \( p \) in \( s\.list() \) do
25. send a “reject” message to \( t; \)
26. \( s\.list() = s\.list() \setminus t; \)

27. Algorithm ends: \( \text{end} \leftarrow \text{true} \) when no new messages are issued;

22 \( \sim \) 23. Then \( s \) will shrink its preference list by removing all the UEs appearing after \( p \) in the list, and it will send a “reject” message to each of the removed UE’s. Note this may also includes its previously partnered UE. The algorithm ends when no messages can be issued by any entity in the network.

The above algorithm is the distributed version of the Gale-Shapley algorithm proposed in [91]. It can be proved that it finishes in \( O(N_P + N_S) \) iterations and results in a stable matching which is optimal for the UEs [70,91]. Note that since the UEs are the owners of the channel and can proactively lease the channel for higher utilities, the result of our
mechanism is therefore desirable. Also, a more complicated egalitarian stable matching can be found as in [94].

3.5 Performance Evaluation

In this section, we evaluate the performance of MIMO-CCRN under different system parameters.

3.5.1 Simulations for the Case of One Primary Link

First we consider a topology where there exists one primary link and \( K = |S| \) secondary pairs. Each SU in the system (ST and SR) is equipped with two antennas. The distance between PT and PR, each with a single antenna, is 100 meters. To reduce the number of model parameters, we adopt a simple geometrical model where the SUs are all located at approximately the same distance \( d (0 < d < 100) \) from the PT and \( 100 - d \) from the PR as in [33][73]. The channel model is decomposed into a large-scale component with path loss exponent \( \eta = 3 \), and a small-scale Rayleigh fading component with \( \sigma = 1 \). For the secondary network, we assume the average channel gain between \( ST_i \) and \( ST_j/SR_j \) is \( 1/40^\eta \) for \( i = j \) and \( 1/120^\eta \) for \( i \neq j \). According to [95], the bandwidth of the primary spectrum is set to 6 MHz and thermal noise level is -129.5 dBm. The transmitting power of PT, \( P_p \), is fixed to a value such that the average SNR of the primary link is 0 dB. The cost per unit transmission energy is \( w = 10 \). Each point in the figures is averaged over 300 independent frame durations.

Figure 3.8 shows the average utility of the primary link, in terms of the throughput under different schemes versus the distance \( d \). We use “MIMO-CCRN: OPT” to represent the performance of MIMO-CCRN through exhaustive search of the best relay set. To reduce the complexity, we restrict our search to the sets \( S(i,j), 0 \leq i, j \leq K \). \( S(i,j) \) is the relay set constructed by including the top \( i \) SRs and top \( j \) STs with the best downlink channel gains \( |h_{0r}|^2 \) in \( \mathcal{R} \). The intuition is to greedily enhance the capacity of Phase One, which is the bottleneck of the cooperative communication. The performance given by this heuristic
algorithm is denoted as “MIMO-CCRN: Heuristic”. In addition, we use “Single-Antenna CCRN” to denote the CCRN scheme proposed in [33], which also aims at maximizing the PU’s throughput with each SU equipped with a single antenna. For fair comparison, both schemes adopt the same settings, and the power budget for SUs are all set to $P_{\text{max}}^r = P_r$.

We assume the target rate $R_k^{(s)}$ of each secondary pair is 5 Mb/s. The number of secondary pairs is $K = 8$. Finally “Direct Transmission” gives the primary link’s throughput without SU cooperation.

From the figure, we can see that by exploiting the SU cooperation with MIMO capability, MIMO-CCRN significantly improves the throughput of the primary link by up to 110% compared with direct transmission. Also, MIMO-CCRN outperforms the Single-Antenna CCRN by up to 75%. This is due to: (i) Secondary relays equipped with multiple antennas achieve stronger beam-forming in receiving and forwarding the PU’s data; (ii) By exploiting the spatial domain, no dedicated fraction of time is allocated to the SUs, thus the overhead for the PU’s transmission is reduced. Moreover the primary link’s throughput reaches a peak when $d$ is around 30, which is the location that best balances the uplink/downlink capacity. Also our heuristic algorithm gives a fairly acceptable performance, which is about 85% of the optimal PU’s throughput for most of the points.
Figure 3.9: Impact of the number of relays.

Figure 3.9 shows the impact of the number of secondary pairs $K$ on the primary link’s throughput. We can see when $K$ increases, the primary link’s throughput improves. For example, when $d = 30$, the PU’s throughput increases from 9.36 Mb/s to 10.83 Mb/s as $K$ changes from 4 to 8. This is because when $K$ becomes larger, there will be more choices for the primary link to choose the secondary relay sets $S_1$ and $S_2$, thus potentially finding better relay sets to enhance its throughput.

Table 3.1 depicts the relationship between the optimal parameter $\alpha^*$ and the distance $d$. We set $K = 8$. As $d$ increases, the downlink capacity from PT to the secondary relays tends to decrease while the uplink capacity from the relays to the PR tends to increase. Therefore, to achieve the optimal overall throughput, the time duration needed for the first phase will increase, which leads to a larger $\alpha^*$.

Table 3.1: Optimal $\alpha^*$ versus the distance $d$.

<table>
<thead>
<tr>
<th>d</th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
<th>50</th>
<th>60</th>
<th>70</th>
<th>80</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha^*$</td>
<td>0.21</td>
<td>0.32</td>
<td>0.43</td>
<td>0.53</td>
<td>0.64</td>
<td>0.73</td>
<td>0.81</td>
<td>0.83</td>
</tr>
</tbody>
</table>

Figure 3.10 shows the comparison of the utilities achieved by the PUs and SUs in
MIMO-CCRN and Single-Antenna CCRN for different $K$. We set $R_k^{(s)}$ of each secondary pair to be 10 Mb/s in MIMO-CCRN, and the power budget $P_r^{max}$ of SUs in both schemes to be the same. For the SUs, since the settings for all the secondary pairs are homogeneous, we measure the utility achieved by a single pair averaged over 300 frame durations. The left part of the figure displays the SU results. As we can see, in both the $d = 30$ and $d = 50$ cases, MIMO-CCRN gives a larger utility to the SUs than Single-Antenna CCRN. The reason is that by exploiting the additional spatial domain in MIMO-CCRN, SUs can be scheduled to transmit their own data throughout the frame duration, instead of being confined to a dedicated time fraction in Single-Antenna CCRN scheme. Moreover, as $K$ increases, more secondary pairs are participating in the competition for the spectrum, which results in a decrease in average utility gained by a single secondary pair. This phenomenon conforms to basic economic principles. The right part of Figure 3.10 shows the corresponding average utility achieved by the primary link for each simulation point, which has similar trends with what is observed in Figure 3.8. When $K$ increases, the utility gained by the PU increases as in Figure 3.9. The results verifies that MIMO-CCRN realizes a stronger win-win situation for PUs and SUs. When $K$ increases, the utility gained by the primary link increases as in Figure 3.9.

Figure 3.11 provides a closer look at the utilities achieved by the SUs. There are $K = 8$ secondary pairs in the network. We modify the topology so that they are placed at $d = 10, 20, 30, \ldots, 80$, one pair for each location. Figure 3.11 shows the average utility achieved by each pair for different secondary transmission rates $R_k^{(s)}$ of SUs. We can see that when $R_k^{(s)}$ increases from 5 Mb/s to 10 Mb/s, the average utility of SUs increases. This is because the gain obtained by the larger data rate overrides the cost of the energy used for transmission, as in Eqn. (3.17). Moreover according to Eqn. (3.23) and (3.24), the relaying power $P_k$ for each secondary pair will increase, which potentially improves the uplink capacity from the relays to the PR. Therefore when $R_k^{(s)}$ is 10 Mb/s, the secondary pairs closer to the PT possess higher priority to be chosen as relays to improve the overall achievable rate of the cooperative communication. We can see that they enjoy higher average utilities compared to the SUs further from the PT.
3.5.2 Simulations for the MIMO Femtocell Case

Previous single primary link simulations reveal some insights for the impact of the system parameters on the performance of MIMO-CCRN. In reality, the network scale would be much larger. To validate MIMO-CCRN in practice, we conduct a large-scale evaluation in this section.

We assume a cellular network. There exist $N_P = 20$ UEs randomly distributed within a semi-circle with a radius of 500 meters centered at a BS. The number of secondary networks $N_S$ is varying from 20 to 100 and all the SAPs are assumed to be randomly distributed within the semi-circle. Each SN still includes one SU$_u$ for uplink and one SU$_d$ for downlink. They are randomly placed in a circle centered at the SAP, with a radius of 50 meters. All the channel parameters remain the same as in the previous section. We assume each UE operates on $|C_{UE}| = 5$ subchannels, each with a bandwidth of 1 MHz. Each SAP owns one free subchannel. The transmission power of SU$_u$ is $0.2P_p$. $w_u$ and $w_d$ are both set to 0.5. The utility functions of the UEs and SNs are defined as their achievable data rates. This setting well reflects our motivating example shown in Figure 3.6 and mimics half a single macrocell. Algorithm 1 is used to find a stable matching in the network. The

![Graph showing utilities achieved by PUs and SU](image-url)

Figure 3.10: The utilities achieved by PUs and SUs.
Figure 3.11: Average utility of the secondary pairs.

The presented results are averages over 20 random topologies. For each topology, we run 300 frame durations before collecting the data. The 90% confidence interval is shown for each point.

Figure 3.12 and 3.13 show the system performance with different numbers of SAPs. We can see from the figures that MIMO-CCRN achieves the highest utilities for both UEs and SNs. For example, when $N_S = 60$, MIMO-CCRN out-performs the Single Antenna and Direct Transmission schemes by 16.6% and 27.8%, respectively, in terms of the average utility of the UEs. It also results in 52.5% and 71.2% larger average utility for the SNs. These results validate that our MIMO-CCRN framework can achieve big “win-win” gains in practical scenarios as compared to the traditional schemes. In Figure 3.12, we can see that the average utility improves monotonically as $N_S$ increases from 20 to 60, since more SAPs in the neighborhood allow more opportunities for the UEs to find suitable cooperative relays.

In Figure 3.13, we can see that as $N_S$ increases, the average utility of the SNs decreases in all the schemes. For example, the utilities drops by 30.3%, 31.9% and 8.9% in both MIMO-CCRN and Single Antenna schemes, respectively, as $N_S$ increases from 20 to 100. As more SNs exist in the area, their competition for the limited spectrum resource owned by the UEs becomes more intense. This leads to a decrease in the utility gain that each of the SNs can achieve.
3.6 Conclusion

This work represents a novel design called MIMO-CCRN to leverage MIMO in cooperative cognitive radio networks. It allows the SUs to cooperatively relay the traffic for the PUs while simultaneously transmitting their own traffic. We design the system archi-
tecture in two network scenarios by considering both the temporal and spatial domains to improve spectrum efficiency. The system-wide competition is carefully considered. Simulation results show that both the primary and the secondary network achieve higher utility in MIMO-CCRN than in the conventional schemes.
Chapter 4

Cooperation among Wireless Service Providers

In this chapter, we study the cooperation among wireless service providers. We investigate the potential benefits of sharing infrastructure and spectrum among two cellular operators. Specifically, we provide a multi-cell analytical model using stochastic geometry to identify the performance gain under different sharing strategies, which gives tractable and accurate results. To validate the performance using a realistic setting, we conduct extensive simulations for a multi-cell OFDMA system using real base station locations. Both analytical and simulation results show that even a simple cooperation strategy with modest changes to the existing network can increase capacity by $30\% \sim 120\%$ under typical conditions. The technical content of this chapter appeared in our previous works [96, 97].

4.1 Background and Motivation

The recent commercialization and popularization of 3G networks have significantly enhanced the mobile users’ capability of making use of ubiquitous data access. However, as users “mobilize” all of their communication activities, from streaming video to cloud computing tasks, cellular operators are facing a severe data capacity crisis. A recent study of Cisco predicts that mobile data traffic will skyrocket by a factor between $25\sim50X$ by 2015 [1]. Such explosive data traffic growth will overload the cellular infrastructure and result in either poor or expensive service to the subscribers. To address this challenge, the
most straightforward solutions, such as adding cells and increasing spectrum have become either expensive or inefficient options.

Cooperation among operators is a solution which has not drawn much attention from researchers. As cellular operators are devoting most of the efforts in expanding their respective networks due to the urgency of the data capacity crisis and market competition, it has become apparent that there exist huge variations in spectrum usage, channel quality and coverage in different operators’ networks [37]. Such diversity generates plenty of cooperation opportunities, which can be exploited to improve their network performance, spectrum efficiency and user experience.

One motivating example is shown in Figure 4.1. Two operators provide cellular service in one area, which are marked in black and grey, respectively. We assume their base stations (BSs) are deployed following a hexagonal layout (solid and dotted lines) and are interweaved. Without cooperation, the two operators will serve their own users solely using their own spectrum bands. The edge users, such as those at points B, D and F in the left black cell, will experience bad channel condition and strong inter-cell interference. However, if the two operators cooperate by allowing their users to flexibly connect to any BS that is closest to them, users at B, D and F will be served by the grey BSs and have excellent channels to the BSs. Generally, the users in the area $\triangle ABC$, $\triangle ECD$ and $\triangle EFA$ in the original black cell will enjoy performance gains, who are mostly edge users with low data rates. A similar effect happens to the users of the grey operator as well, such as those in the
area $\triangle AOC$. As we can see, simple cooperation improves the capacity of both operators and results in a win-win situation. Depending on the levels of synergy, the operators can share each other’s infrastructure, spectrum, or even merge into a single operator. Interestingly, in the AT&T’s proposed acquisition of T-Mobile, one of the reasons cited is the resulting increase in capacity [98].

In this chapter, we focus on studying the benefits brought by different levels of cooperation among cellular operators. We are interested in answering the following questions: In what ways can the operators cooperate? How much performance improvement can their cooperation bring? Only a few recent papers address this issue [99] [100] [36]. In [99], Chandramani et al. optimize the operators’ aggregate payoffs given BS locations and user rates, and use game theory to discuss how to share the profits. Supratim et al. in [100] develop a user choice algorithm with network information provided by the operators. Peng et al. in [36] focus on spectrum-based cooperation, and form a group bargaining model based on the demand on each BS.

However, all the above papers require cross-layer optimization based on the instantaneous network status. They do not characterize the long-term network capacity improvement for a large-scale multi-cell deployment. Our research fills this gap and provides a tractable and accurate model for quantifying the gains under different cooperation strategies. To validate our model, we have collected a set of real BS location data and provide performance results using real-world layouts under an Orthogonal Frequency Division Multiple Access (OFDMA) system. This aspect also has not yet been examined before. We expect our research to give an insight into the benefit of cooperation and guide operators considering a range of cooperation options. Specifically the contribution of this chapter is two-fold:

- We present an analytical model for identifying the performance gain from cooperation among cellular operators. It provides a tractable and reasonably accurate model for average user rate/throughput under a multi-cell environment.
- We perform extensive simulations using real BS locations and OFDMA resource
allocation algorithms. This validates the advantage of operator cooperation in a realistic setting.

4.2 Analytical Modeling

In this section, we derive an analytical model to evaluate the network performance when the cellular operators cooperate. We consider two strategies, with different levels of cooperation, that we envisage the cellular operators may adopt:

1) **FLEXROAM** (short for “flexible roaming”): the cellular operators allow their users to freely connect to BS of either operator that provides the best signal strength. It will require an update in the signaling protocols to facilitate this.

2) **MERGER**: in addition to FLEXROAM, all the BSs of an operator can reuse the spectrum of its cooperating operators. This refers to an additional sharing of spectrum, which may be the result of a full merger of the two operators, but does not preclude a mutual business agreement short of a full merger.

Note that FLEXROAM and MERGER can also be offered to users by a mobile virtual network operator, or MVNO, which can purchase wholesale use of the infrastructure from the two operators. We further use **NOCOOP** to refer to the scheme when the two operators do not cooperate.

Cellular networks are traditionally modeled by assuming the BSs are placed following a hexagonal layout [2]. However, these models have long suffered from being both intractable and highly idealized. Recently, a general model based on stochastic geometry was proposed in [101], which provides tractable ways to evaluate network performance considering inter-cell interference and fading. Our analytical modeling follows the methods in [101], and provides tractable results of the performance under the various modes of operators’ cooperation.

We will mainly focus on the user’s **average ergodic rate** and **average throughput**, which are important metrics for benchmarking the cellular system. We assume there are two operators, OP\(_1\) and OP\(_2\), coexisting in the network. Our models can be easily extended
to include more than two operators, and to derive other performance metrics, such as the distribution of user’s Signal-to-Interference-plus-Noise Ratio (SINR).

We assume the locations of BSs of OP₁ and OP₂ follow independent Poisson point processes Φ₁ and Φ₂ with densities λ₁ and λ₂, respectively. The model drawn from such a random deployment is shown to be about as accurate as the standard grid model, compared to an actual cellular network [101]. Further we use \( W_i \) to denote channel bandwidths used by the BSs of OP\(_i\), \( i \in \{1, 2\} \). We consider the downlink performance of the system. Without losing generality, we assume a typical user is located at the origin. If it is connected to the BS \( b_0 \) of OP\(_i\), its SINR can be expressed as:

\[
\text{SINR}_i = \frac{P_t h_{b_0} r_{b_0}^{-\alpha}}{N_0 W_i + P_t \sum_{b \in \Phi_i \setminus b_0} g_b r_b^{-\alpha}},
\]

where \( P_t \) is the fixed transmission power of all the BSs, \( N_0 \) is the noise power density. The distance between the user and BS \( b_0 \) is \( r_{b_0} \), and the channel fading is \( h_{b_0} \). The user’s distances to other interfering BSs \( b \) in \( \Phi_i \) are \( r_b \) and the corresponding channel fadings are \( g_b \). \( \alpha \) is the path loss exponent. We assume all the fadings are Rayleigh fading with mean 1.

### 4.2.1 Average User Ergodic Rate

Following the proofs in [101], when the two operators do not cooperate and run their service as is, the average ergodic rate in the downlink for a typical user of OP\(_i\)’s is given by:

\[
R_{\text{NOCOOP}}(W_i, \lambda_i) = \mathbb{E}[\ln(1 + \text{SINR}_i)] = \int_{r > 0} e^{-\pi \lambda_i r^2} \int_{t > 0} e^{-\frac{N_0 W_i r^{\alpha} (e^t - 1)}{r_t}} \frac{1}{r_t} F(\lambda_i) dt 2\pi \lambda_i r dr,
\]

where

\[
F(\lambda_i) = \exp(-\pi \lambda_i r^2 (e^t - 1)^{2/\alpha}) \int_{(e^t - 1)^{-2/\alpha}}^{\infty} \frac{1}{1 + x^{\alpha/2}} dx.
\]

Eq. (4.2) and (4.3) can be computed numerically and are used for performance comparisons in this work. Next we examine the performance under cooperation strategies. When OP₁ and OP₂ share infrastructure following the FLEXROAM strategy, we have the
Theorem 1: Under FLEXROAM strategy, for a typical user, the average ergodic rate is given by:

$$R_{FLEXROAM}(W, \lambda) = \frac{\lambda_1}{\lambda_1 + \lambda_2} R_1(W_1, \lambda) + \frac{\lambda_2}{\lambda_1 + \lambda_2} R_2(W_2, \lambda),$$

where

$$R_i(W_i, \lambda) = W_i \int_{r>0} e^{-\pi(\lambda_1+\lambda_2)r^2} \int_{t>0} e^{-\frac{N_0 W_i r^2 (e^t-1)}{h_2}} \cdot F(\lambda_i) d\pi(\lambda_1 + \lambda_2)r dr.$$  (4.4)

and $F(\lambda_i)$ is given by Eq. (3). $W = \{W_1, W_2\}$ and $\lambda = \{\lambda_1, \lambda_2\}$ are the vectors for bandwidths and BS densities.

Proof: We use $r$ to denote the distance between the user and its associated BS. Under FLEXROAM, the user will connect to the closest BS from either of the operators that gives the best average signal strength. The union of all the BSs is still a Poisson point process $\Phi$ with density $\lambda_1 + \lambda_2$. The cumulative density function (CDF) of the distance $r$ is then:

$$F_r(R) = Pr[r \leq R] = 1 - Pr[r > R] = 1 - Pr[\text{no BS from any operator closer than } R] = 1 - e^{-(\lambda_1+\lambda_2)\pi R^2}. $$  (4.5)

Thus the probability density function (PDF) of $r$ is:

$$f_r(r) = \frac{dF_r(r)}{dr} = e^{-(\lambda_1+\lambda_2)\pi r^2}2\pi(\lambda_1 + \lambda_2)r. $$  (4.6)

Moreover, we have

$$R_{FLEXROAM}(W, \lambda) = \int_{r>0} f_r(r) E_{h,g,\Phi}(\ln(1 + \text{SINR})) dr. $$  (4.7)

We further use $X_i = 1$ to represent the event that the user is associated to the BS of OP$_i$. Note $X_i \in \{0, 1\}, \forall i = 1, 2$ and $X_1 + X_2 = 1$. According to the rule of total probability,

$$E_{h,g,\Phi}(\ln(1 + \text{SINR})) = \sum_{i=1,2} E_{h,g,\Phi,i}(\ln(1 + \text{SINR}_i))Pr[X_i = 1], $$  (4.8)
wherein, $\text{SINR}_i$ is given by Eq. (1). Using $D_i$ to represent the user’s distance to the closest BS from OP, in a manner similar to Eq. (7), the CDF of $D_i$ is $f_{D_i}(r) = e^{-\lambda_i \pi r^2} 2\pi \lambda r$. We have:

$$\Pr[X_1 = 1] = \Pr[D_1 < D_2] = \int_0^\infty \int_r^\infty e^{-\lambda_2 \pi r_2^2} 2\pi \lambda_2 r_2 e^{-\lambda_1 \pi r_1^2} 2\pi \lambda_1 r_1 dr_2 dr_1$$

$$= \frac{\lambda_1}{(\lambda_1 + \lambda_2)}.$$  \hspace{1cm} (4.9)

Similarly, we have $\Pr[X_2 = 1] = \frac{\lambda_2}{(\lambda_1 + \lambda_2)}$. Plugging Eq. (4.8) and (4.9) into (4.7), and following the rest of the steps as in Appendix B of [101] completes the proof.

We next discuss the average ergodic rates when the MERGER strategy is employed. Under this scheme, each BS reuses the spectrum owned by the other operators, e.g., they can operate on the whole spectrum $W_1 + W_2$. As a result, the BSs from both operators interfere with each other. We then have the following corollaries from Eq. (2) and (3):

**Corollary 1.** Under the MERGER strategy, for a typical user, the average ergodic rate is:

$$R_{\text{MERGER}}(W, \lambda) = R_{\text{NOCOOP}}(W_1 + W_2, \lambda_1 + \lambda_2).$$  \hspace{1cm} (4.10)

**Proof:** This is same as the single operator case with BSs of density $\lambda_1 + \lambda_2$ operate on bandwidth $W_1 + W_2$. We then have the corollary following Eq. (2) and (3).

4.2.2 Average User Throughput

We will now determine the average throughput for a typical user under FLEXROAM and MERGER. We use $\eta_i$ to denote the subscriber density of OP, $i \in \{1, 2\}$. Generally, subscriber density is far larger than the macrocell BS density, thus we assume $\eta_i \geq \lambda_j, \forall i, j \in \{1, 2\}$.

According to the law of large numbers, in NOCOOP, if the two operators serve their users independently, the number of users in each cell of OP will be approximately $\frac{\eta_i S}{\lambda_i S} =$
\[ \eta \frac{i}{\lambda}, \text{ where } S \text{ is the total area under consideration.} \]  
We further assume that the scheduling decisions on the BSs ensure \textit{proportional fairness}, which is a common design objective of all current and next generation cellular systems. With proportional fairness, the average throughput a user of OP \( i \) under NOCOOP scheme is:

\[
Th_{\text{NOCOOP}}(W_i, \lambda_i, \eta_i) = R_{\text{NOCOOP}}(W_i, \lambda_i) \cdot \frac{\lambda_i}{\eta_i}, \quad (4.11)
\]

Moreover, under FLEXROAM and MERGER strategy, the average number of users in each cell becomes \( \eta_1 + \eta_2 \). Therefore, the average throughput for a typical user is:

\[
Th_{\text{FLEXROAM}}(W, \lambda, \eta) = R_{\text{FLEXROAM}}(W, \lambda) \frac{\lambda_1 + \lambda_2}{\eta_1 + \eta_2}
\]

\[ = \frac{1}{\eta_1 + \eta_2} \left( \lambda_1 R_1(\lambda_1, \lambda_2) + \lambda_2 R_2(\lambda_1, \lambda_2) \right). \quad (4.12)\]

where \( \eta = \{\eta_1, \eta_2\} \). Finally, under the MERGER strategy, the average throughput for a user is:

\[
Th_{\text{MERGER}}(W, \lambda, \eta) = R_{\text{MERGER}}(W, \lambda) \frac{\lambda_1 + \lambda_2}{\eta_1 + \eta_2}. \quad (4.13)
\]

Compared to the traditional hexagonal model that requires complex system-level Monte Carlo simulations, our analytical model can be computed using a numerical tool in a short time.

### 4.2.3 Special Case: Cooperation among Same-Size Operators

To elicit some insights from our model, we consider a special case where two same-size operators cooperate. Specifically, we assume \( W_1 = W_2 = W, \lambda_1 = \lambda_2 = \lambda, \eta_1 = \eta_2 = \eta \). Moreover, as cellular networks are typically interference-limited, we assume there is no noise, i.e., \( N_0 \rightarrow 0 \).

Under these assumptions, Eq. (3.11) to (3.13) can be greatly simplified. For example,

\[
Th_{\text{FLEXROAM}}(W, \lambda, \eta) = \frac{\lambda}{\eta} \cdot R_1(W_1, \lambda),
\]

\[ = \frac{\lambda W}{\eta} \int_{t>0} e^{-2\pi \lambda t^2} \int_{r>0} F(\lambda) dt 4\pi \lambda r dr, \quad (4.14)\]

\[ = \frac{\lambda W}{\eta} \int_{t>0} \frac{2}{2 + G(t)} dt. \quad (4.15)\]
where $G(t) = (e^t - 1)^{2/\alpha} \int_{(e^t - 1)^{-2/\alpha}}^\infty \frac{1}{1+x^{\alpha/2}} dx$. Eq. (4.14) and (4.15) can be derived by letting $v = \pi \lambda r^2$ and integrate over $v$. Similarly, we can derive the average user throughput under MERGE and NOCOOP as:

$$T_{h,\text{MERGER}}(W, \lambda, \eta) = \frac{\lambda W}{\eta} \int_{t>0} \frac{2}{1 + G(t)} dt, \quad (4.16)$$

$$T_{h,\text{NOCOOP}}(W, \lambda, \eta) = \frac{\lambda W}{\eta} \int_{t>0} \frac{1}{1 + G(t)} dt. \quad (4.17)$$

From Eq. (4.16) to (4.17), we can see that MERGER exactly doubles the average user throughput that is achieved by NOCOOP. This is due to the doubling of the spectrum that can be used by both operators. Further, to evaluate FLEXROAM, we numerically compute Eq. (4.15) and (4.17) for two typical values of $\alpha$: $\alpha = 3.5$ and $4$. We found that FLEXROAM improves the average user throughput by 44.6% when $\alpha = 4$ and 45.9% when $\alpha = 3.5$. Note that more results will be presented in Section 4.4.1.

### 4.3 An OFDMA-based Multi-Cell System

In the previous section, we described an analytical model for evaluating the network performance under different cooperation strategies of the operators. However, the model simplifies the real system in three aspects:

- it assumes Poisson random BS deployment;
- it uses the entire spectrum without considering subchannelization;
- it assumes perfect resource allocation following the proportional fairness criterion.

Since the constraints above may significantly change performance, it is still useful to validate network performance results for a practical multi-cell system under the various cooperation strategies. To evaluate the benefits of operator cooperation in realistic system, we plan to combine real BS location data and practical user-subchannel scheduling algorithms. As OFDMA is widely adopted by all the next-generation cellular standards (WiMAX, LTE) to divide the spectrum into subchannels, in this section, we present a low complexity OFDMA
A typical OFDMA system partitions the radio resource in both frequency and time domains as shown in Figure 4.2. The total bandwidth is divided into $C$ orthogonal subchannels, and one frame is divided into $T$ time slots. A subchannel and time slot combination, termed a *tile*, is the minimum unit for resource allocation. In one BS, a tile can only be assigned to one user to avoid intra-cell interference. For each BS $b$, we use $\mathcal{M}_b$ to denote the set of users associated with it. Thus there are totally $|\mathcal{M}_b|$ users in this cell. For a specific BS, $\mathcal{M}_b$ may vary depending on the different cooperation strategies the operators use. Specifically, in NOCOOP, $\mathcal{M}_b$ are the users from the same operator of $b$ that are connected to $b$, whereas in FLEXROAM and MERGER, $\mathcal{M}_b$ include the users from the other operators that treat $b$ as the best BS. In addition, the number of subchannels $C$ may also vary according to the cooperation strategy. It will increase in MERGER as each BS reuses the spectrum owned by the other operators.

In the multi-cell scenario, neighboring cells may reuse the same tile depending on the inter-cell interference. As a result, the optimal multi-cell resource allocation problem is combinatoric and generally regarded as an NP-hard problem. Only heuristic methods exist so far [102] [103] [104]. Next, we present a centralized greedy algorithm to achieve a sub-optimal solution with low computational complexity. It assumes the users’ channel state information is available at a Radio Network Controller (RNC), which coordinates
all the BSs. Note that since our purpose is to quantify the gain brought by the operator cooperation, designing more practical algorithms is beyond our scope. Our algorithm is generalized from the ones in [102] [103]. It runs for each frame and aims to maximize the total utilities of all the users throughout all the cells in a frame. The utility of user $i$ is defined as $U(R_i) = \log(R_i)$, with $R_i$ being its data rate in the current frame. Using log function as the utility has been proved to be equivalent to finding a proportional fair solution [105]. Assuming there are $B$ BSs and $N$ users, the complexity of the algorithm is $O(TCBN)$.

As shown in Algorithm 1, Line 1 initializes the matrices that contain the allocation results. Each entry $R_{i,t}$ in $R$ records the aggregate data rate of user $i$ up to the slot $t$. $X$ and $Y$ are boolean matrices. The entry $Y_{b,c,t} = 1$ in $Y$ if BS $b$ is transmitting on subchannel $c$ in slot $t$. The entry $X_{i,c,t} = 1$ in $X$ indicates that the user $i$ is receiving on subchannel $c$ in slot $t$. The subchannels are allocated sequentially (Line 3), and for each one, we check each BS for whether it should transmit on this subchannel in the current slot. To maintain fairness, the BSs with fewer number of assigned subchannels are given higher priority (Lines 6~7).

<table>
<thead>
<tr>
<th>Algorithm 1: Multi-Cell Resource Allocation Algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 $R_{K\times T} = 0$; $Y_{B\times C\times F} = 0$; $X_{K\times C\times T} = 0$;</td>
</tr>
<tr>
<td>2 for $t = 1 : T$ do</td>
</tr>
<tr>
<td>3 for $c = 1 : C$ do</td>
</tr>
<tr>
<td>4 $\Omega \leftarrow$ set of all the BSs;</td>
</tr>
<tr>
<td>5 while $\Omega \neq \emptyset$ do</td>
</tr>
<tr>
<td>6 $b^* \leftarrow$ BS with the least number of assigned channels in $\Omega$;</td>
</tr>
<tr>
<td>7 $\Omega \leftarrow \Omega \setminus b^*$;</td>
</tr>
<tr>
<td>8 $Y' \leftarrow$ set $Y_{b^*,c,t} = 1$ in $Y$;</td>
</tr>
<tr>
<td>9 foreach $m \in M_{b^*}$ do</td>
</tr>
<tr>
<td>10 $\Delta U_m = U_{gain,m} - U_{loss,m}$;</td>
</tr>
<tr>
<td>11 $m^* \leftarrow \arg\max_m \Delta U_m$;</td>
</tr>
<tr>
<td>12 if $\Delta U_{m^*} &gt; 0$ then</td>
</tr>
<tr>
<td>13 $X \leftarrow$ set $X_{m^*,c,t} = 1$ in $X$;</td>
</tr>
<tr>
<td>14 $R \leftarrow \forall i, X_{i,c,t} = 1$, set</td>
</tr>
<tr>
<td>15 $R_{i,t} = R_{i,t} - r_{i,c,t}(Y') + r_{i,c,t}(Y)$ in $R$;</td>
</tr>
<tr>
<td>16 $Y \leftarrow Y'$;</td>
</tr>
</tbody>
</table>
Our algorithm is marginal utility driven. For the current BS $b^*$, Lines 9~13 picks the user with the maximal marginal utility $\Delta U$. For each user $m$ associated with $b^*$, $\Delta U_m$ is defined as the difference between the utility increase $U_{gain,m}$ by scheduling user $m$ and the total utility loss of other users on the same subchannel $U_{loss,m}$ due to the increased interference (Line 10). $U_{gain,m}$ and $U_{loss,m}$ can be evaluated by:

$$U_{gain,m} = U[R_{m,t} + r_{m,c,t}(Y')] - U(R_{m,t})$$  

$$U_{loss,m} = \sum_{i,X_i,c,t=1}^{i} \{U(R_{i,t}) - U[R_{i,t} - r_{m,c,t}(Y) + r_{m,c,t}(Y')]\}$$

$Y'$ is the matrix assuming BS $b^*$ is transmitting on subchannel $c$ in slot $t$. The function $r_{m,c,t}(Y)$ is the rate of user $m$ on subchannel $c$ in slot $t$ based on the allocation matrix $Y$. We assume the transmit power is equally split across the subchannels on each BS. Considering the inter-cell interference, $r_{m,c,t}(Y)$ is defined as:

$$r_{m,c,t}(Y) = W_C \log_2(1 + \frac{P_t/C \cdot h_{b_0,m}^c \cdot d_{b_0,m}^{-\alpha}}{N_0 W_C + \sum_{b \neq b_0,Y_{b,c,t} = 1} h_{b,m}^c \cdot d_{b,m}^{-\alpha}}),$$

where $W_C = W/C$ is the bandwidth of a subchannel, $b_0$ is the BS user $m$ associates to. $h_{b,m}^c$ is the subchannel fading between BS $b$ and user $m$ on subchannel $c$, and $d_{b,m}$ is the distance between BS $b$ and user $m$. Finally, Lines 12~15 update the matrices if the maximal marginal utility is positive.

### 4.4 Performance Evaluation

In this section, we assess the performance of our proposed cooperation strategies through two sets of simulations. In the first set, we numerically compute the expressions given by the analytical model described in Section 4.2. To further show the results under practical settings, in the second set, we conduct simulations for the OFDMA system with the multi-cell resource allocation algorithm shown in Section 4.3.

We have obtained precise coordinates for BSs from two major operators over a large suburban area near Washington D.C.. The $20 \times 20$ km area we chose is shown in Figure 4.3.
There are 16 BSs from the one operator and 13 BSs from the other. We mark the BSs using the colors red and blue to distinguish them. The location information will be used for our simulation.

4.4.1 Numerical Evaluations

Here we numerically compute the results of our analytical model when different cooperation strategies are employed by the operators. We will focus on user’s average throughput as it is a critical metric for evaluating cellular service. It is important to show how much improvement there is through operator cooperation, and how such an improvement varies when the two operators have different BS densities, user densities and bandwidths.

We set the value of \( \lambda_1 \) in Eq. (12)~(14) as the red operator’s BS density, e.g., \( \lambda_1 = \frac{16}{400000000} = 4 \times 10^{-8} \text{ m}^{-2} \). \( \lambda_1 \) will be fixed in all the scenarios while \( \lambda_2 \) is left as an adjustable parameter. According to IEEE 802.16m evaluation methodology document [106], we set the transmission power of the BSs as \( Pt = 46 \text{ dBm} \), noise power density as \( N_0 = -174 \text{ dBm/Hz} \) and the path loss exponent \( \alpha = 3.76 \).
The Impact of BS Density

In this scenario, we set both OP_1 and OP_2’s bandwidths to $W_1 = W_2 = 10$ MHz according to [106]. And we let their respective user densities be $\frac{n_1}{\lambda_1} = \frac{n_2}{\lambda_2} = 100$. As a result, their spectrum resource and average number of users per cell are identical. Figure 4.4 shows the results by changing the BS density ratio $\lambda_2 / \lambda_1$. As $\lambda_2$ changes from $0.2 \lambda_1$ to $2 \lambda_1$, the performance of OP_2 under NOCOOP only slightly increases, since $\lambda$ does not
influence the signal-to-interference ratio and the system is interference limited. We can easily see cooperation brings significant benefits to both operators. When $\frac{\lambda_2}{\lambda_1} = 1$, the two operators have 193.3 kb/s average user throughput without cooperation. However, if they cooperate, FLEXROAM and MERGER strategies improve the average user throughput to 281.0 kb/s and 387.4 kb/s respectively, which is equivalent to a 45.4% and 100.4% increase, respectively. As $\lambda_2$ keeps increasing, the performance of FLEXROAM will slightly drop since the opportunistic diversity is reduced. However, FLEXROAM and MERGER still achieve a large performance gain for both operators within a wide range of BS density ratios.

The Impact of User Density

To explore the impact of user densities on the network performance of cooperation, we fix $\lambda_2 = \lambda_1 = 4 \times 10^{-8}$ m$^{-2}$ and $W_2 = W_1 = 10$ MHz. Then we let $\eta_1 = 100\lambda_1$ and vary $\eta_2$ from 0.2$\eta_1$ to 2$\eta_1$. This represents different user/BS ratio for OP$_2$, e.g., when $\eta_2 = 0.2\eta_1 = 20\lambda_2$, on average there are only 20 users per cell. Figure 4.5 shows the results. As we can see, as $\eta_2$ increases, under NOCOOP, the average user throughput of OP$_2$ drops drastically when more users are sharing the radio resource. In contrast, under
FLEXROAM and MERGER the performance degrades gradually since users of OP\(_2\) can still opportunistically connect to the BSs of OP\(_1\). Moreover, when OP\(_1\) and OP\(_2\)'s user densities are different, they achieve different performance gains. For instance when \(\frac{\eta_2}{\eta_1} = 0.8\), under NOCOOP, OP\(_1\) and OP\(_2\) have 193.26 kb/s and 241.58 kb/s average user throughput. When FLEXROAM is adopted, the performance jumps to 312.21 kb/s, which gives 61.5% and 29.2% gains to OP\(_1\) and OP\(_2\), respectively (marked in the circles in the figure). Generally, FLEXROAM results in a win-win situation when \(\frac{\eta_2}{\eta_1}\) ranges from around 0.5 to 2.0. MERGER has a larger win-win range, which is from 0.37 to beyond 2.0. When the difference between \(\eta_1\) and \(\eta_2\) is too large, cooperation may generate negative effects to one operator. For example, when \(\frac{\eta_2}{\eta_1} = 0.4\), OP\(_2\) loses performance under FLEXROAM. However, since we just are studying a simple roaming policy, such negative effects are expected to be avoided by adopting more advanced policies and incentive mechanisms.

**The Impact of BS Bandwidth**

In this scenario, we fix \(\eta_1 = \eta_2 = 100\lambda_1 = 100\lambda_2\). We fix \(W_1 = 10\) MHz, and change the ratio \(W_2/W_1\). As we can see from Figure 4.6, the performance of FLEXROAM, MERGER and OP\(_2\) under NOCOOP increases almost linearly as the total bandwidth increases. When \(\frac{W_2}{W_1} > 0.4\), FLEXROAM provides better performance for both OP\(_1\) and OP\(_2\). It enables more gain for the operator with less spectrum. MERGER achieves a larger gain for all the points in the figure.

**4.4.2 Performance with Real BS Locations**

To validate the numerical results in practice, we conducted the OFDMA-based simulations with real BS locations. We consider two cellular networks with BSs placed as in Figure 4.3. Mobile devices are deployed uniformly in the 20 km \(\times\) 20 km area with a density of 100 users per cell for each operator. The system is assumed to be coordinated by the OFDMA resource allocation algorithm we proposed in Sec. III. Based on the IEEE 802.16m evaluation methodology document [106], the main system parameters are summa-
Table 4.1: OFDMA System Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of subchannels</td>
<td>32</td>
</tr>
<tr>
<td>Number of slots per frame</td>
<td>60</td>
</tr>
<tr>
<td>BS Transmit Power</td>
<td>46 dBm</td>
</tr>
<tr>
<td>Noise power spectrum density $N_0$</td>
<td>-174 dBm/Hz</td>
</tr>
<tr>
<td>Channel bandwidth</td>
<td>10 MHz</td>
</tr>
</tbody>
</table>

For the channel gain, we model the fast fading component as Rayleigh fading with mean 1. We also considered shadow fading and model it as a log-normal random variable with a standard deviation of 8 dB. The overall effect of antenna gain, cable and penetration loss and noise figure is set to 0 dB. Finally, the path loss at the distance $d$ is modeled as $L(d)_{dB} = 17.39 + 3.76 \log_{10} d$.

Depending on the different cooperation strategies, the user set of each BS may be different. Under NOCOOP and FLEXROAM, each operator still aims to optimize the total utility of all the users connected to its own BSs, while under MERGER, we optimize the global utility of the two operators. We ran each simulation for 30 consecutive frames and we averaged the results of 5 simulation runs. Further, we plug the density values $\lambda_1 = 16/4000000000 = 4 \times 10^{-8}$, $\lambda_2 = 13/400000000 = 3.25 \times 10^{-8}$, $\eta_1 = 100\lambda_1$, $\eta_2 = 100\lambda_2$ into our previous analytical model to resolve the numerical results for comparison purposes.

Figure 4.7 shows the results. Under NOCOOP, the performance of OP$_2$ is worse than
OP_1 since it has fewer BSs and the resulting average signal is weaker. Overall, the average throughput of all the users in the system under NOCOOP is around 263.8 kb/s. Cooperation substantially improves the performance of both the operators the real setting. Specifically, the user’s average throughput increases to 386.24 kb/s in FLEXROAM and 519.31 kb/s in MERGER. These give 34.4% and 80.75% improvement to OP_1 and 68.21% and 126.16% improvement to OP_2 respectively. Moreover, our analytical model accurately captures this trend and provides very close improvement predictions in terms of percentage. The OFDMA system generally offers better performance than the numerical results since it better exploits the user/frequency selectivity. Also, the analytical model is pessimistic since it assumes the random deployment of BSs, which is discussed in [101].

To further scrutinize the gain for the users, we plot the CDF of the average throughput of the users in one simulation run. Figure 4.8 shows the results in the domain of 0 ~ 200 kb/s. For NOCOOP, we use the statistics from all the users in the system. The curves validate the improvement result from cooperation. For example, the median value of user’s average throughput increases by around 40% and 100% under FLEXROAM and MERGER compared to NOCOOP.
4.5 Conclusion

This work investigates the potential benefits of cooperation among cellular operators. Using stochastic geometry, we provide a tractable analytical model to derive the average user rate and throughput under two cooperation strategies. With real base station locations, extensive simulations over a multi-cell OFDMA system further validate the performance improvement. We show that even a simple cooperation policy with modest changes to existing networks can achieve around 30%~120% capacity gains under typical conditions.
Chapter 5

Conclusion and Future Work

5.1 Summary of Major Contributions

In this thesis, we consider the designs that exploit the benefits of cooperation in wireless networks. Specifically, three forms of cooperation are discussed, including multi-hop cooperative networking, physical layer cooperative communications and the cooperation among wireless service providers. Within each domain, we propose novel architectures and analysis to help the system design and improve the performance. The interplay of different layers in the protocol stack is carefully considered and our approach answers the questions of how to utilize the advanced physical layer technique and how to improve the application side performance, within the context of cooperation.

First, as mobile video service is expected to become a popular application for cellular network operators while still suffers from some efficiencies, we propose to integrate multi-hop cooperative networking and scalable video coding to improve its performance. In our solution, called SV-BCMCS, video is encoded into one base layer and multiple enhancement layers using SVC. Different layers are broadcast with different modulation and coding profile to cover viewers at different ranges. As it is a common fact that most of the cell phones are equipped with multiple interfaces, we allow viewers to forward enhancement layers to each other using short-hop and high-rate ad-hoc connections. Based on the above framework, we study the optimal resource allocation problem and design an efficient helper discovery scheme. Further, a multi-hop video content forwarding scheme is
proposed to exploit the broadcast nature of ad-hoc transmissions and eliminate redundant video relays. We also develop an analytical model to study the expected gain of cooperative multi-hop forwarding under random user distribution in a cellular cell. Extensive video trace driven simulations using OPNET are conducted with three representative video sequences, which show that SV-BCMCS significantly improves the video quality perceived by users in a practical hybrid cellular network scenario.

Second, we study the influence of PHY layer cooperative communications to the cognitive radio networks. Conventional cooperative cognitive radio networks (CCRNs) operate in temporal domain only, which limits the performance of both primary users (PUs) and secondary users (SUs). We extend the CCRN scheme into both spatial and temporal domains by exploiting the multiple antennas equipped on the SUs. We study the feasibility of cooperative communications with MIMO technique by designing correct beamforming vectors. Our proposed framework, called MIMO-CCRN, enables the SUs to utilize MIMO to cooperatively relay the traffic for the PUs, while concurrently accessing the same channel to transmit their own traffic. Further, we consider typical network scenarios. For the case of a general PU link cooperating with multiple SUs, we formulate an optimization model based on Stackelberg game to maximize the utilities of PUs and SUs. For the case of a practical cellular network with multiple MIMO-empowered femtocells, we provide an algorithm to find a stable matching of the PUs and SUs. It is shown in our simulations that leveraging MIMO spatial cooperation results in a large “win-win” situation for PUs and SUs.

Third, as cellular operators are facing a severe data capacity crisis due to the explosive demand of the users to “mobilize” all of their communication activities, we investigate new methods based on the cooperation among operators to solve this problem. We consider two main cooperation options, FLEXROAM and MERGER. An analytical model based on stochastic geometry is presented for identifying the performance gain under these two strategies. It is tractable and reasonably accurate to reflect the average user rate/throughput under multi-cell environment. Moreover, for more realistic settings, we consider real BS locations and OFDMA resource allocation algorithms and conduct extensive Monte-Carlo
simulations. Both analytical and simulation results show that even a simple cooperation policy with modest changes to existing networks can achieve around 30% to 120% capacity gains per customer under typical conditions.

5.2 Future Work

5.2.1 Content Delivery over Hybrid Cellular/Ad-hoc Networks

Beyond SV-BCMCS architecture, several relevant issues are left as our future work:

Multiple Cells: The technique of soft-combining across multiple cells for EV-DO BCMCS is proved to benefit the edge users by enhancing their data throughput and transmission reliability. Currently we didn’t integrate the soft-combining into our optimal resource allocation but only consider the single-cell case. This is left as our future work. With soft-combining enabled, the edge users will experience higher data rate, while the users in the middle of the cell can also have their multicast rate improved due to SV-BCMCS, which leads to a more fair performance.

Interference in Ad-hoc Network: Interference in the main factor limiting the throughput of multi-hop wireless ad-hoc network. Even without collision, simultaneous transmissions in the same channel will interfere with each other, resulting in the degradation of the link rates. The OPNET-based simulations we conducted already include the interference effect. While simulations give a fairly good performance, it still trails the theoretical bound assuming perfect ad-hoc transmission, due to interference. One of our future work is to optimize the multicast tree in ad-hoc relay network. Such multicast tree guarantees each user still receiving the video layers it should receive decided in resource allocation, while minimizing the network interference.

Moreover, as cellular networks are currently suffering from severe data capacity crisis, multi-hop cooperation over ad-hoc networks is considered an important method for performance improvement. One of our future approach is to leverage the ad-hoc network for content recovery. In detail, if a group of closed-by users are interested in a common
content, the BS can do a single broadcast. The content is assumed to be encoded using net-
work coding [107] and cached on each user [108][109], and different users have different
receiving packet error probabilities. Then they can use WiFi links to recovery the content.
Another future direction is to investigate the benefits of using Delay Tolerant Networks
(DTNs) [7] for content delivery. DTN is a kind of network to distribute the data based on
the mobility of the users. It is already shown recently in [110] that DTN can be leveraged
for cellular traffic offloading. However, how to best utilize DTN in the context of cellular
network is remained an interesting problem to study. Besides the above, we are also trying
to build up testbeds for realistic video streaming in the hybrid network scenario.

5.2.2 Cooperative Cognitive Radio Networks

Our proposed scheme MIMO-CCRN is the first work studying how to utilize MIMO
cooperation in cognitive radio networks. In our study, we assume each SU is equipped with
two antennas to simply our beam-forming designs and game theory analysis, and we only
discussed the feasibility of beyond two antenna systems. In the future, we plan to carefully
consider the Degree-of-Freedom for beamforming designs when both PUs and SUs have
multiple antennas. Also, currently we assume as long as PUs and SUs are cooperating,
all the spectrum of the PU is allocated to SU. A more general scenario includes a partial
spectrum leasing. For example, when OFDMA is present, a subset of subcarriers can be
assigned to SUs while the PU can still use the left ones for transmission. How much
bandwidth is partially leased is determined by the benefits generated via cooperation.

5.2.3 Cooperation among Wireless Service Providers

Cooperation among wireless service providers is a broad area in which, our study is
mainly at an initiating stage. However, it already shows that simple cooperation policies
can achieve huge capacity gains and mutual benefits to all the service providers. In the
future, one of our plans is to investigate the feasibility and possible changes to the existing
infrastructure to enable the cooperation. This needs us to look at into more details of the
core network and the signaling. As all cellular operators are converging to LTE, we expect only modest changes are enough to support operator cooperation. Another direction is designing more practical cooperation strategies beyond FLEXROAM and MERGER. It is validated that cooperation can achieve an overall capacity benefits. However, when the participating operators have different size of BSs, users and spectrum, how to fairly share the resulting benefit is a problem worth investigating. As cooperators are selfish entities and have both cooperation and competition, flexible pricing mechanisms may be necessary to direct the user to a correct cell selection. The user demand, cell congestion level and operator policies also need to take into consideration. In addition, besides the cooperation among cellular operators, the cooperation among other wireless service providers is also an interesting problem. For example, there can be cooperations among private WiFi hotspots, or the WiFi network and cellular network. As heterogeneous network is becoming a trend and mobile terminals are generally equipped with multiple interfaces, cooperation is a way to best utilize the current resource to achieve a good user experience. For example, an initial study on using femtocells with relays to offload traffic is shown in [111].
Bibliography


[98] “Acquisition of T-Mobile USA, Inc. by AT&T Inc. - Description of Transaction, Public Interest Showing and Related Demonstrations,” *Filed with the Federal Communications Commission*, April 2011.


