DEVICE-TO-DEVICE ASSISTED COOPERATIVE COMMUNICATIONS

by

Seyed Hossein Seyedmehdi

A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy
Graduate Department of Electrical and Computer Engineering
University of Toronto

Copyright © 2014 by Seyed Hossein Seyedmehdi
Abstract

Device-to-Device Assisted Cooperative Communications

Seyed Hossein Seyedmehdi
Doctor of Philosophy
Graduate Department of Electrical and Computer Engineering
University of Toronto
2014

In a cooperative communications scenario, wireless devices assist each other in their transmissions or receptions. This cooperation has been shown to improve the communication reliability while increasing the efficiency of the spectrum usage.

In this thesis, we study how a device, which may be idle otherwise, can be leveraged to increase the spectral efficiency of the wireless system. First, we consider a two-user case where a secondary user perceives the environment and adapts its behaviour in a way that it does not degrade the performance of a primary user (also known as the “cognitive radio”). Moreover, we consider a multi-user case where single antenna devices can be clustered together to virtually form a more sophisticated device that can perform beamforming.

In the two-user cognitive radio, under the half-duplex assumption and collision avoidance model, the primary and secondary instantaneous data rates are derived as a function of the power allocation factor by the secondary user. Also, the performance of various cooperative schemes including decode-and-forward and amplify-and-forward is investigated.

The two-user case is further extended to the multi-user case where the problems of clustering (of the source device with a subset of perhaps idle devices) and beamforming are separated and studied. For the clustering problem, a simple and agnostic greedy algorithm is shown to perform optimal. Also, two beamforming techniques are proposed: firstly, a selfish beamforming using the feedback channel, and secondly a coordinated beamforming (also known as interference alignment) using the reciprocal channel.
Acknowledgements

By starting a PhD, each student embarks upon a journey to an uncharted territory hoping to discover new grounds. Like an explorer, what this student discovers is greatly influenced by the people he encounters. Now that I have reached a destination after a rewarding graduate study period, I am grateful to all those who helped and inspired me to move forward in the right direction.

Firstly, I must thank my PhD advisor, Professor Ben Liang, for his support. Furthermore, I am grateful to the members of my supervisory committee, Professor Alberto Leon-Garcia and Professor Baochun Li; my examination committee members, Professor Shahrokh Valaee, Professor Elvino Sousa, and Professor Alberto Leon-Garcia; and the thesis external appraiser Professor Dongmei Zhao, for their invaluable feedback.

A thank you to all my friends, fellow PhD students, and officemates (Amir Tavakoli, Andrei, Dafna, Siavash, Mohammad Mahanta, Mohammad Norouzi, Saeid Rezaei, Mahdi, Amin, Yunfeng Lin, Amir, Ali Kalantarian, Kaveh, Alireza, Ahmad, Nima, Samira, Amir, Padina, Yana) who shared this experience with me and helped me to stay motivated. An additional thank to my fellow explorer and dear old friend, Amir Aghaei, for many enlightening discussions and things I learned from him.

I would also like to thank the knowledgeable people at Ericsson Canada (Adrien, Peter, Karl, Vijay, Norbert, Talha, Mitch) for giving me the learning opportunity during my internship year, and especially, I am deeply grateful to my manager, Ron Casselman, and my mentor, Gary Boudreau, PhD., for their remarkable support and empathy during the most difficult phase of my life.

Most important, my special gratitude goes to my parents, my sisters and my brother for their constant support, continuous encouragement, unconditional love, and sharing their invaluable life experience. Without them, this thesis would not have been completed.
Contents

1 Introduction 1
   1.1 Motivation .................................................. 2
   1.2 Contributions .............................................. 6
   1.3 Thesis Outline ............................................. 7
   1.4 Notations and Definitions ................................. 8

2 Background and Related Literature 10
   2.1 Cognitive Radio ........................................... 10
      2.1.1 Full-Duplex Cognitive Radio ....................... 11
      2.1.2 Half-Duplex Cognitive Radio ...................... 12
   2.2 Device-to-Device Assisted Virtual MIMO ............... 15
   2.3 Iterative Beamforming and Clustering for Virtual MIMO 17

3 Half-Duplex Cognitive Radio with Collision Avoidance 21
   3.1 System Model and Problem Statement .................... 22
   3.2 Achievable Rates of the Cooperative CR ................. 25
      3.2.1 Decode-and-Forward Strategies ................... 25
      3.2.2 Amplify-and-Forward Strategies ................. 30
   3.3 Throughput Analysis of the Cooperative CR ............. 33
   3.4 Numerical Illustrations .................................. 36
      3.4.1 Data Rate for a Moving Cooperative CR .......... 36
3.4.2 Effect of Channel Parameters on Throughput ........................................ 37

3.5 Chapter Summary ...................................................................................... 38

4 Half-Duplex Cognitive Radio ........................................................................ 43

4.1 System Model ............................................................................................ 43

4.2 Physical Layer Cooperation Scheme .......................................................... 44

4.3 Power Allocation Algorithms ...................................................................... 48

4.3.1 Selfless Power Allocation ........................................................................ 48

4.3.2 Greedy Power Allocation ........................................................................ 48

4.3.3 Numerical Results and Discussions ........................................................ 50

4.4 Chapter Summary ...................................................................................... 50

5 Clustering for Device-to-Device Assisted Virtual MIMO ............................... 51

5.1 System Model and Problem Statement ...................................................... 52

5.1.1 Distribution of UEs and Scheduling Model ........................................... 52

5.1.2 Structure of the VMIMO and Corresponding Rates .............................. 53

5.1.3 Statement of the Problem ...................................................................... 56

5.2 An Approximation Method for the Precoding Problem ............................ 57

5.3 Clustering for a Single Source .................................................................... 60

5.3.1 An Upper Bound on the Average Spectral Efficiency .......................... 60

5.3.2 An Optimal Relay Selection Algorithm ............................................. 63

5.4 Clustering for Multiple Sources ................................................................. 66

5.5 Numerical Results ..................................................................................... 68

5.5.1 Simulation Setup and Parameters ......................................................... 68

5.5.2 Discussion ............................................................................................. 70

5.6 Chapter Summary ...................................................................................... 73

6 Iterative Beamforming for D2D Assisted VMIMO ....................................... 74

6.1 System Model and Problem Statement ..................................................... 75
B.6 Proof of Lemma 1 ......................................................... 111
B.7 Proof of Proposition 1 .................................................. 113
B.8 Proof of Theorem 9 ...................................................... 114
B.9 A Constructive Protocole to Achieve $\tilde{r}_i$ in Definition 6.1.1 ....... 115
B.10 Proof of Proposition 2 .................................................. 116
B.11 Derivations of the Precoding Vectors with Per-Antenna Power Constraints . . . . 118

C Full-Duplex Causal Cognitive Radio 120

C.1 Definitions and Channel Modelling ......................................... 120
C.2 Main Results ........................................................................ 121
  C.2.1 An Achievable Rate Region for the Causal CR ......................... 121
  C.2.2 Relationship between Theorem 11 and Existing Results .............. 124
  C.2.3 On The MIMO Broadcast Channel ...................................... 125
C.3 The Gaussian Channel ............................................................ 126
C.4 Proof of Theorem 11 ............................................................. 128
  C.4.1 Codebook Generation ...................................................... 128
  C.4.2 Encoding ....................................................................... 130
  C.4.3 Decoding ....................................................................... 131
  C.4.4 Error Probability Analysis ............................................... 132

Bibliography 132
List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>List of cooperative strategies studied in this chapter.</td>
<td>33</td>
</tr>
<tr>
<td>5.1</td>
<td>Description of the legends in Fig. 5.3</td>
<td>65</td>
</tr>
<tr>
<td>5.2</td>
<td>Improvement in the harmonic mean of the spectral efficiency for the coverage threshold $\text{SINR}_{\text{eff}} = -10 \text{ dB}$</td>
<td>72</td>
</tr>
<tr>
<td>5.3</td>
<td>Increase in the average consumed energy per bit $E_b (J/b)$ due to the VMIMO setup</td>
<td>72</td>
</tr>
<tr>
<td>6.1</td>
<td>Definition of some of the system variables</td>
<td>77</td>
</tr>
<tr>
<td>6.2</td>
<td>Simulation Parameters</td>
<td>87</td>
</tr>
</tbody>
</table>
List of Figures

3.1 Mnemonic channel model for the cooperative cognitive radio. ............... 22
3.2 Demonstration of the channel use over time by the PU and the SU. ............ 34
3.3 Geographical locations of the users and data rates achievable for the SU when
the ST travels. ........................................................................................................ 39
3.4 Throughput illustration for the DF strategy with SIC at the PR for different
values of $P_2$. ........................................................................................................ 40
3.5 Throughput versus $P_2$ for six different strategies discussed in Sec. 3.2.1 and
Sec. 3.2.2. ........................................................................................................... 41
3.6 Throughput versus the channel gain $c$ for six different strategies discussed in
Sec. 3.2.1 and Sec. 3.2.2. .................................................................................... 42
3.7 Throughput versus the channel gain $c$ for six different strategies discussed in
Sec. 3.2.1 and Sec. 3.2.2. .................................................................................... 42
4.1 A diagram representation for the 4-node cooperative cognitive radio. ........... 45
4.2 The actions taken by each node during two consecutive blocks. .................. 46
4.3 Comparison between achievable rates in DDF in [1], full duplex DF in [2], and
the half duplex DF in Theorem 7 for $\beta = 1$. .................................................. 47
4.4 Throughput of the SU vs. the duty cycle $\xi$ for different $P_2$’s. .................... 49
4.5 Power allocation factors vs. the duty cycle $\xi$ for the greedy power allocation. 49
Glossary

**AF** amplify-and-forward: a relaying technique in which the relay only amplifies the received signal from the source and forwards it through which the noise at the relay is also amplified and forwarded. 6, 15, 16, 21, 25, 30, 33, 36, 37, 95, 96

**AP** access point: any point of contact that is connected to the back hull network including base stations (BSs) and enhanced node-Bs (eNBs). 8, 20, 51–55, 60, 63, 64, 68, 69, 75–79, 81–84, 89, 94, 95, 109, 113

**ASE** average spectral efficiency: average number of information bits that can be conveyed free of error to the destination per unit of time and bandwidth. The unit of the average spectral efficiency is bits per second per hertz (bps/Hz). 33, 38, 44, 48, 50, 57, 61–63, 65, 93

**CCR** causal cognitive radio: a cognitive radio in which the secondary user acquires the knowledge about the message of the primary user through the channel in causal manner (also known as the cooperative cognitive radio). 11, 12, 120, 126, 128, 133

**CoCR** cooperative cognitive radio: a cognitive radio scenario in which the secondary user cooperate with the primary user to be granted the access to the channel (analogous to the causal cognitive radio). 3, 6, 7, 12, 13, 21, 22, 25, 37, 38, 43, 50, 93

**CR** cognitive radio: a concept in which a secondary user perceives the environment and adapt its communication in a way that it can innocuously co-exists with a primary user. 1, 2
CSI  channel state information. 7, 8, 19, 20, 75, 81, 88, 89, 91, 92, 95

D2D  device-to-device. 5, 8, 73, 74, 89, 95

DDF  dynamic decode-and-forward: a decode-and-forward relaying scheme first introduced in [1]. 12, 43, 46, 47

DF  decode-and-forward: a relaying technique in which the relay decodes the message of the source and then forwards it. 6, 15, 16, 21, 25, 27, 29, 30, 33, 36, 38, 40, 46, 47, 51, 67, 78, 93, 96, 108, 124

DMC  discrete memoryless channel: a communications channel in which the messages are chosen from a set of discrete alphabets. 97, 120, 122

DPC  dirty paper coding: a coding method that can cancel the effect of the interference when it is known at the receiver [4]. 12, 13, 43, 45, 93, 120, 127, 128

FD  full-duplex: a transceiver technology in which a wireless device is capable of both receiving and transmitting on the same spectral band at the same time. 11, 129, 14, 120

GMIMO-BC  Gaussian MIMO broadcast channel. 125, 128, 133

HD  half-duplex: a transceiver technology in which a wireless device is either capable of receiving or transmitting on the same spectral band at any particular time. 4, 7, 11, 13, 21, 43, 93, 94

IC  interference channel: a four-node communication channel comprising two co-channel transmitters and corresponding receivers. 11, 128

ICC  interference channel with conferencing: a four-node communicate channel that comprises two transmitters and two respective receivers where transmitters are capable of overhearing the channel [5]. 11, 12, 124, 125, 128, 133
**LTE**  long-term evolution: a standard for wireless communication of high-speed data for mobile phones and data terminals. 2, 17, 74, 75

**MAC**  medium access control. 2, 4, 5

**max-SINR**  max-SINR: a beamforming algorithm in which beamforming vectors are the MMSE decoders of the reciprocal channel. 7, 18, 19, 86, 88, 90, 92

**MIMO** multi input multi output: any communication systems in which the receiver is equipped with multiple receive antennas, the similarly, the transmitter is equipped with multiple transmit antennas. 1, 2, 5, 16, 18, 19, 51, 74, 125, 126

**MMSE**  minimum mean squared error. 18, 19, 55, 79, 97

**MSE**  mean squared error. 18, 19, 79, 80, 86, 87, 94, 100

**MWSMSE**  minimum weighted sum mean squared error: a beamforming algorithm that tries to minimize the sum of weighted MSEs. 7, 19, 86, 88, 90

**PHY**  physical layer. 4, 5

**PM**  proximity measure: a measure to determine how suitable two objects are to be clustered together. 7, 20, 82, 84, 86, 88, 89, 91, 95

**PPP**  Poisson point process: a probabilistic model for spatial distribution of users in which the number of users in a given area follows a Poisson distribution. 6, 7, 51, 52

**PR**  primary receiver: defined in the context of the cognitive radio, the primary receiver is the receiver node of the primary user. 13, 15, 22, 23, 27, 33, 86, 40, 102, 104

**PT**  primary transmitter: defined in the context of the cognitive radio, the primary transmitter is the transmitter of the primary user. 11, 14, 15, 22, 23, 25, 28, 30, 32, 34, 36, 37, 39, 102, 104
**PU** primary user: defined in the context of the cognitive radio, a primary user is the often the user that is licensed to utilize the bandwidth. [1] [3] [4] [7] [10] [14] [21] [22] [24] [25] [32] [35] [37] [38] [40] [43] [46] [48] [51] [93] [95] [104] [106] [108] [120]

**RAT** radio access technology: refers mostly to the different radio access standardizations including LTE and IEEE802.11. [8]

**RAT** sub 1 first radio access technology. [75] [87] [115]

**RAT** sub 2 second radio access technology. [74] [76] [87] [89] [115]

**RB** resource block: a portion of time and frequency that is usually assigned to a user. [2]

**SDR** semidefinite relaxation: an approximation method for solving non-convex quadratically constrained quadratic optimization problems. [6] [7] [52] [57] [58]

**SIC** successive interference cancellation: a receiver structure in which the messages are decoded individually, and after each message is decoded, it is subtracted from the received signal, thus reducing the interference for the subsequent decodings. [6] [15] [21] [27] [33] [36] [38] [40] [93]

**SINR** signal to interference and noise ratio. [3] [18] [23] [57] [58] [68] [69] [94] [101]

**SINR** sub eff effective signal to interference and noise ratio: an alternative measure of spectral efficiency defined in [5.23] [69] [72]

**SNR** signal to noise ratio. [12] [60] [65] [94]

**SR** secondary receiver: defined in the context of the cognitive radio, the secondary receiver is the receiver node of the secondary user. [13] [15] [22] [24] [27] [33] [37] [39] [103] [104]

**ST** secondary transmitter: defined in the context of the cognitive radio, the secondary transmitter is the transmitter of the secondary user. [11] [14] [15] [21] [22] [24] [28] [30] [34] [36] [39] [104]
**SU** secondary user: defined in the context of the cognitive radio, a secondary user is often the user that is not licensed to utilize the bandwidth unless it does not harm the primary user’s performance. [1, 3, 4, 6, 7, 10, 15, 21, 22, 24, 25, 32, 40, 43–46, 48, 51, 93, 95, 103, 104, 107, 108, 120]

**UE** user equipment: it can include any mobile device such as cellphones, sensors, or PDAs. [2, 4, 8, 10, 19, 20, 51, 54, 56, 59, 79, 81, 84, 86, 88, 92, 94, 96, 109, 110, 113–116]

**VD** virtual device: a collection of devices that are grouped together and assist each other in the transmission or reception. [5, 8, 19, 74–79, 81, 86, 88, 89, 95, 116]

**VMIMO** virtual MIMO: a MIMO system in which multi antennas where antennas are located on different physical devices as opposed to the conventional multi antenna systems with all the transmit antennas on the same device. [2, 5, 8, 17, 19, 20, 51, 53, 54, 57, 59, 62, 65, 69, 73, 89, 95]
Chapter 1

Introduction

Wireless spectrum is an expensive and scarce commodity, and there is a rapidly growing demand for it. For instance, wireless internet traffic in North America is expected to grow more than 41-fold from 2011 to 2016. In addition, mobile data usage doubled in 2012. Smartphones use 50 times the amount of spectrum as a basic feature phone, while tablets use 120 times the amount of spectrum [6]. These observations imply that either the spectrum available to expand mobile services must increase, or the current available but limited spectrum resource must be utilized more efficiently.

Over the past decade, there have been numerous efforts to utilize the spectrum resource in the most efficient way. Specifically, there have been two concepts that tend to show more potential than others to achieve this goal. The first promising concept was the notion of a software-defined radio or cognitive radio (CR) as introduced in [7]. In the CR networks, a secondary user (SU) is capable of perceiving the environment and adapting its communication in such a way which minimizes the harm to the primary user (PU). The second promising technology, originally introduced more than three decades ago [8], is the use of multiple transmit and receive antennas, also known as multi-input-multi-output (MIMO) in cellular systems. Nowadays, the use of MIMO has been widely accepted as a promising technique to improve the spectral efficiency of wireless communication systems, and it is supported by various current
standards, including the long term evolution (LTE) and IEEE 802.11. However, in practice, due to the limited size of a user equipment (UE), usually only one antenna can fit inside each UE. Recently, to achieve the MIMO predicted gains for these single antenna devices, the concept of virtual MIMO (VMIMO) has been proposed as a promising future trend [9]. In a VMIMO system, multiple (single-antenna) UEs can be grouped together to create a virtual multi antenna device. In this thesis, we explore both the CR and VMIMO as means of enhancing the spectral efficiency.

1.1 Motivation

Nowadays, most current wireless standards and implementations have a conservative approach to distribute resources. For instance, in the LTE standard, resource blocks (RBs) are exclusively assigned to individual UEs in each cell, and UEs are not allowed to use any other RBs other than the granted ones. This rigidity increases the wireless system reliability by controlling collisions and interference (unlike the competitive approach in IEEE 802.11), thus ensuring a certain level of quality of service. Another reason for adopting this approach is to simplify the medium access control (MAC) protocols. That is, this straightforward approach significantly reduces the implementation complexity in terms of the number of man-hours required to develop the scheduler software. The disadvantage of this approach, however, is that only a limited number of UEs can be serviced at any given time, thus leaving most of them idle specially in a dense deployment of UEs. Since UEs are resources that have the potential of improving the performance of the system, attempting to minimize the number of idle ones is desired.

Relaying is one of the most intuitive ways of utilizing idle devices. A relay is a wireless node (or UE) that forwards either a decoded or an amplified version of its received signal. This received signal at the relay is often from another wireless node or UE that is termed as the source in the wireless communication jargon. As the result of this relay action, the performance of the source is improved. The reason for this improvement is that the due to
co-transmission of the relay, the received signal to interference and noise ratio (SINR) at the destination is increased and therefore, the spectral efficiency (which is a function of the SINR) is also increased. Most of the time the relay has no message of its own and only forwards a modified version of its received signal. However, if this relay has its own message and intended destination, it can cooperate with the source to convey the source message while mitigating the induced interference to the source. This form of cooperation is known in the literature as cooperative cognitive radio (CoCR). In CoCR, the source is the PU which is licensed to utilize the spectrum resources, and a SU acting as a relay with its own message that can transmit only if it does not degrade the performance of the PU. To achieve this goal, the SU perceives the environment and adapts its way of communication.

Throughout the first half of this thesis\(^1\) the following fundamental question is tried to be addressed: \textit{what is the maximum spectral efficiency that can be achieved for a CoCR?}\(^2\) It should be noted that in the CoCR considered in this thesis, there are two users: a PU and a SU. The PU acts as the source and the SU acts as the relay with a message of its own. The SU can adopt different strategies through which the spectral efficiency of both PU and the SU may vary. As a simple example, if the SU is completely selfless meaning that it allocates all of its resources to forward the message of the PU, its spectral efficiency will be zero (since it conveys no information of itself) whereas the spectral efficiency of the PU is maximized. Therefore, the cooperation strategy (with the resource allocation being one part of it) often plays an important role to determine the achievable spectral efficiency. This fact leads to another question: \textit{what is the best cooperation strategy?}

Addressing the last question without considering any practical constraint can lead to a complex solution that is not implementable\(^3\). Therefore, in Chapters 3 and 4 solutions are investigated within the boundaries of practical constraints. The first example of such constraints is...
the collision avoidance in order to simplify the receiver design. Under this assumption, PU and SU avoid concurrent transmission. This constraint is, however, relaxed later in Chapter 4. The second of these assumptions is the half-duplex (HD) assumption. Based on this assumption, a UE can either transmit or receive at any given frequency at any given time.

From another perspective to the problem of improving the spectrum utilization, it must be noted that the number of UEs is predicted to increase by an order of magnitude over the next few years. This growth will naturally cause a similar increase in the density of UEs (number of UEs per unit area). While it might be perceived as only hindering the communication by increasing the interference level, this abundance of UEs can actually be utilized to improve the spectral efficiency through cooperative communication. However, given the high density of UEs, an important problem needs to be addressed: How to select the best subset of available UEs as relays for the source? To be practical, any solution to this problem must accommodate two properties: firstly, it must not incur any significant communication overhead; secondly, it must be fast.

For an efficient relay selection algorithm, three elements must be jointly considered: physical (PHY) layer, MAC layer, and the algorithm run-time. The final objective of the relay selection is to improve a performance aspect of the PHY layer including spectral efficiency and/or reliability. However, this improvement is often achieved at the cost of more complicated MAC layer protocols, extensive communication overheads, or a latent selection algorithm with a run-time that grows very fast with the number of available UEs.

Ideally, the selection algorithm must consider all the available UEs (the search space) and select the best subset of them as the actual relays. Obviously, as the size of the search space increases, the required time to perform such an algorithm increases. It is noteworthy that this run-time is additional to the time for the overhead signalling communications and tends to be neglected when the search space is small. Nonetheless, the behaviour of the algorithm as the search space grows is of significant importance. For instance, an exhaustive search (which has

---

4Both these constraints are relaxed in Appendix C.
an exponential run-time) may seem to be a viable option for a small number of UEs, but it quickly becomes computationally intractable as the cardinality of the search space increases. In Chapter 5, a selection algorithm is proposed that is optimal from a PHY layer perspective and has a linear run-time (the best that can be achieved) and incurs a low overhead (MAC layer).

In addition to relaying, devices can cooperate with each other through other means. One of these means is the additional antennas (on relaying UEs) that can be exploited to form a VMIMO system. In a conventional MIMO system multiple antennas are located on the same physical device. However, with the size of UEs constantly shrinking, inside each UE usually only one transmit antenna can fit specially in low cost or legacy devices. These single antenna devices can be clustered together to form a virtual multiple antenna device (device-to-device (D2D) assisted VMIMO). A new challenge arise in this realm: a distributed beamforming method must be developed.

There are two ways of developing a distributed beamforming method when multiple transmitting virtual devices (VDs) are considered: non-coordinated and coordinated. In the non-coordinated method, the beamforming weights for each VD is determined regardless of other VDs and there is usually a feedback channel involved. In a coordinated method (e.g., interference alignment), each VD selects its beamforming weights such that its footprint interference to other VDs is mapped to a space orthogonal to their signal space. The coordinated method can potentially achieve higher performance because of the extra degree of freedom. However, almost all coordinate beamforming methods are iterative, and therefore, it might take a long time to converge. On the contrary, the non-coordinated algorithm determines the beamforming weights in a single shot. In addition to the convergence time, there are other performance trade-offs between the coordinated and non-coordinated methods which need to be investigated. Examples of these trade-offs include computational complexity, feedback requirement, coping with the aging channel, and energy per information bit. The non-coordinated and coord-

---

5 The interference alignment concept is described in details later.
Dominated beamforming are investigated in Chapters 5 and 6 respectively.

1.2 Contributions

This thesis evaluates different aspects of radio access design by studying the potential benefits of cooperation among the nodes in the physical layer. Below the main contributions of this thesis are enclosed:

- **CoCR** with collision avoidance is considered [10] where decode-and-forward (DF) and amplify-and-forward (AF) strategies for the CoCR are investigated as two prevailing strategies. Different decoders with successive interference cancellation (SIC) and without SIC are considered. For each cooperative strategy, analytical closed forms for the optimal power allocation and maximum achievable rate for the SU were obtained.

- An efficient clustering algorithm for the D2D assisted VMIMO is investigated [11] where single antenna UEs are randomly distributed on a 2D plane according to a Poisson point process (PPP) and only a subset of them are sources leaving other idle UEs available to assist them (relays). It is first shown that the NP-hard optimization problem of precoding in this scenario can be approximately solved using semidefinite relaxation (SDR). A special case with a single source is investigated, and an upper bound on the average spectral efficiency of the VMIMO system is analytically derived. Then, an optimal greedy algorithm is proposed that achieves this bound. These results are further exploited to obtain a polynomial time clustering algorithm for the general case with multiple sources. Finally, numerical simulations are performed to compare the performance of the proposed algorithm with that of an exhaustive clustering algorithm, and it shown that these numerical results corroborate the efficiency of the proposed algorithm.

---

6 These results are mainly included in Chapter 3 of this thesis.
7 These results are mainly included in Chapter 5 of this thesis.
• The iterative beamforming and clustering for the D2D assisted VMIMO is investigated in [12]. The contribution of this work is two-fold: firstly, new iterative beamforming updates are proposed that satisfy the per-antenna (or per UE since each UE has a single antenna) power constraints while requiring only local channel state information (CSIs); secondly, a novel proximity measure (PM) based on the relative distance of the UEs is introduced. The results are also compared with the previously known max-SINR and the minimum weighted sum mean squared error (MWSMSE) based updates, and it is shown that the proposed beamforming updates can improve both sum throughput and fairness while most previous methods fail to achieve so.\(^8\)

### 1.3 Thesis Outline

This thesis is outlined as follows. In Chapter 3, the CoCR under a collision model is considered where the SU cooperates with the PU by overhearing the PU and relaying its message in HD mode. The goal of this chapter is to develop cooperation strategies for the CoCR such that the average spectral efficiency (ASE) of the PU is not degraded while that of the SU is maximized. In addition, these results are extended to the case without collision avoidance in Chapter 4 where a simple power allocation scheme is provided.\(^9\) In Chapter 5, the problem of clustering and beamforming for a D2D assisted VMIMO is considered. In this chapter, our goal is to develop an efficient algorithm to cluster each source with a subset of available relays to form a VMIMO system under a limited feedback assumption. Towards this goal, a polynomial time algorithm is proposed to cluster users and perform the beamforming by applying the SDR. The proposed algorithm is shown to be optimal for a single source case in terms of the average spectral efficiency. Also, when UEs are distributed according to a PPP with finite mean, it is shown that the proposed algorithm has linear complexity in the number of available UEs. In

---

\(^8\)These results are mainly included in Chapter 6 of this thesis.

\(^9\)The problem in Chapter 3 is also considered in Appendix C with the difference that in Appendix C, the half-duplex limitation is relaxed and the secondary transmitter is capable of full-duplex transmission.
Chapter 6, we consider the problem of clustering and beamforming for D2D assisted VMIMO. In the considered scenario, there are a number of co-channel UEs and a number of access points (APs) operating under the following assumptions: Each UE is equipped with a single transmit antenna whereas APs have multiple antennas, and UEs are transmitting to the APs (uplink). Single antenna UEs can be clustered together to form a VD. When clustered, UEs can leverage a second radio access technology (RAT) to share their uplink messages. When UEs form a VD, they transmit their messages using a precoder. The objective is to cluster the UEs together and find the precoding weights to maximize a sum-utility metric while minimizing the required feedback for the CSIs.

1.4 Notations and Definitions

Before moving forward, in what follows, the notations and definitions used in this thesis are explained. Capital letters (\(A\)), lower case letters (\(a\)), and calligraphic letters (\(\mathcal{A}\)) denote random variables (RVs), their sample values, and their alphabet sets respectively. A similar convention is used for random vectors and their values. The notation \(X \sim_{\text{i.i.d.}} p(x)\) is used to indicate that the random variable \(X\) is drawn independent and identically distributed (i.i.d.) according to the probability measure \(p(\cdot)\) on \(\mathcal{X}\). The notation of \(X^i_j, i \leq j\) is used to show the vector \((X_i, X_{i+1}, \cdots, X_j)\), and for brevity, \(X^j_i\) is shown by \(X^j\). Typical sets are shown by \(T^{(n)}(X)\), and the corresponding set \(X\) is omitted when it is clear from the context. For two matrices \(A\) and \(B\), \(A \preceq B\) denotes \(B - A\) is positive semidefinite. The terms rate and spectral efficiency are used interchangeably to indicate the average number of bits per second per hertz (bps/Hz) that can be conveyed through the communication channel. The AWGN channel is assumed to have the noise distribution \(\mathcal{CN}(0, \sigma_N^2)\). By definition, \(C(x) = \log_2(1 + x)\). The ceiling of \(x\) is shown by \([x]\). The backslash notation is used to represent the set subtraction, i.e., \(A \setminus B \triangleq A - B\). Vectors are shown by small bold faced letters (e.g., \(h\)), and matrices are shown by capital bold faced letters (e.g., \(H\)). The \(i\)-th element of a vector \(h\) is shown by \(h_i\). The
operator $|\cdot|$ is either the absolute value when its operand is a complex (or real) number or the set cardinality if it operates on sets, and $\|\cdot\|$ represents the norm. $I_N$ represents the unity matrix with dimension $N \times N$, and when it is clear from the context the subscript $N$ is omitted. Also, $(\cdot)^\dagger$ represents conjugate transpose of a matrix, $(\cdot)^T$ represents transpose of a matrix, and $(\cdot)^*$ represents conjugate of a matrix.
Chapter 2

Background and Related Literature

The previous literature related to this thesis can be classified into three categories based on how the cooperative device is employed: firstly, the body of work that consider the cooperative device as a CRs which cooperates with the PU; secondly, the works that consider the idle devices as relays to improve the performance of the sources; thirdly, the works in which all the UEs are sources that can cooperate with each other. In what follows, a detailed description of the related works in each category is provided.

2.1 Cognitive Radio

CR is a promising concept that addresses the need for scaling the capacity of wireless systems. A CR channel [13–16] refers to a communication model, in which two users (a PU and a SU) communicate with their intended receivers simultaneously via a common communication medium (co-channel), and the SU has causal or non-causal knowledge about the message being transmitted by the PU. In this scenario, the SU perceives the environment and dynamically adapts its transmission strategy so that it can innocuously coexist with the PU [16–18]. For this model, several achievable rate regions and a capacity region in the low interference regime have been established in [14, 15, 19, 21].

Most previous models for CRs rely on the critical assumption that the SU has non-causal
knowledge about the message of the \textbf{PU}. In a pioneering work, Devroye et al. [17] studied the data rate of the non-causal \textbf{CR} assuming a genie informs the \textbf{secondary transmitter (ST)} about the message of the \textbf{primary transmitter (PT)} prior to transmission. However, the non-causality assumption is not valid in realistic communication scenarios, where it is infeasible for the \textbf{SU} to non-causally obtain such knowledge. The \textbf{causal cognitive radio (CCR)} model was later developed in [3], where the authors adopted a cooperative strategy for the \textbf{CR} and provided an inner bound on the joint achievable rates of the \textbf{PU} and the \textbf{SU}. In this model, the \textbf{ST} overhears the message of the \textbf{PU} through the channel. Then the \textbf{ST} transmits its own message superimposed with the overheard message incorporating a superposition coding technique. In other words, the \textbf{SU} acts as a relay with a private message. To mitigate the interference effect, it further splits its rate into two parts. At the receivers, it is assumed that arriving messages from different transmitters are perfectly synchronized.

Several \textbf{CR} models that do not require the non-causality assumption, have been proposed and investigated [3, 19]. In these models, the \textbf{SU} either is capable of overhearing the channel and transmission simultaneously (\textbf{full-duplex (FD)}) or is only capable of performing one at a time (\textbf{HD}). In the latter case, the \textbf{SU} only listens to the channel and then transmits. In this dissertation we study both \textbf{FD} and \textbf{HD} \textbf{CRs}.

\subsection{Full-Duplex Cognitive Radio}

The \textbf{FD CCR} was first investigated in [19], where a causal transmission protocol was proposed and an achievable rate region was obtained. Later, an improved rate region was derived in [3] by introducing a cooperative model, in which the \textbf{SU} acts as not only an senders but also a relay for the \textbf{PU}. Most previous works on the \textbf{CCR} study the \textbf{FD CR} (e.g., [16, 17, 22, 23]), and most often, they involve complicated coding schemes.

Casual knowledge for the interference channels have been investigated under the \textbf{interference channel with conferencing (ICC)} model [5]. The \textbf{ICC} refers to an \textbf{interference channel (IC)} in which both transmitters are simultaneously capable of overhearing from the channel.
In other words, there are two feedback channels from the receivers to the corresponding transmitters. An achievable rate region for the ICC is derived in [5] by splitting the message of each transmitter into three parts; cooperative, common, and private where only the cooperative part is sent cooperatively using block Markov encoding. In Section C.2.2, we discuss the potential drawbacks of this method. It is noteworthy that in the more recent works, the CCR is also known as the CoCR since the SU cooperates with the PU.

In Appendix C, a CCR model in which the SU can obtain the message being transmitted by the PU in a causal manner is investigated. Employing a coding strategy that consists of the block Markov superposition, dirty paper coding (DPC), and backward decoding, we obtain an achievable rate region which includes that in [3]. We discuss the case where each user is able to causally obtain knowledge on the message being transmitted by the other user and show that our results can reduce to some previously known results [24–26]. We further provide numerical comparisons between our results and those in [3] and [5].

2.1.2 Half-Duplex Cognitive Radio

There is a major instrumental issue with the FD CR. Due to the physical limitations of the radio frequency receivers-transmitters and the severe signal to noise ratio (SNR) degradation of the received signal when combined with the transmit signal at the same transceiver, the FD transiers are highly impractical. The HD transiers, on the other hand, overcome this problem by either transmitting or receiving the signal.

A HD relay is the basic building block of any HD cooperation scheme in wireless communications. The relaying schemes and performance of HD relays are studied in the seminal works [1, 30, 31]. In the seminal work by Azarian et al. [1], a dynamic decode-and-forward (DDF) scheme is proposed. In this scheme, the relay listens to the source’s transmission for a fraction of time, and as soon as it can decode the message of the source, it commences the cooperative transmission.

In terms of the receiver design, there are two different approaches: first, receivers that
can not decode the colliding messages, and secondly, receivers that are capable of decoding colliding messages. In what follows, the related literature to the HD CR are surveyed, and we consider both receivers.

**Half-Duplex Cognitive Radio Without Collision Avoidance**

The inner bounds on the HD CR was investigated in [32], which provides an achievable rate region seemingly to be intractable. In [33], a cooperation protocol for the CoCR based on the DPC is proposed and its achievable rates are obtained. ( [33, Eq. (17)]. In the derivation of such rates, it is assumed that the SU has full knowledge about the message of the PU through overhearing the channel, and the SU can use DPC to mitigate the interference caused by the PU. However, the transmission needed to convey such information to the SU is not considered (gene aided channel).

**Half-Duplex Cognitive Radio with Collision Avoidance**

The largest inner bound for the HD CR model is derived in [23], but it involves many auxiliary random variables, so that an analytical closed form for the rate of the SU as a function of the rate of the PU seems to be intractable. One implicit assumption in the most works in this area (e.g., [23], [32]) is that two colliding messages at the receiver can be decoded as if they are perfectly synchronized. However, holding this assumption is very cumbersome in reality. Decoding colliding packets are possible, but it requires highly sophisticated receivers and transmitters. For instance, in [34], multi-user techniques for interference cancellation are discussed. In [35], a cross-layer protocol called DAC is developed to circumvent the synchronization requirement. However, these techniques require complicated signal processing method which are difficult to implement.

In Chapter 3, we consider a CoCR model in which the SU can cooperate with the PU in the HD mode. Furthermore, neither the primary receiver (PR) nor the secondary receiver (SR) can decipher two colliding packets. Hence, our physical channel model resembles that in [36].
The goal for the SU is to maintain the quality of service (QoS) of the PU while optimizing its own spectrum utility.

We assume the PT has a set of bits of information to transmit (a file for instance). After incorporating an error correction method, these bits are coded into $B$ blocks or packets to be transmitted at the front end. When the ST cooperates, the PT and the ST transmit in alternating slots, with the ST serving as a relay for the PT by superpositioning a version of the PT's block with its own data block. This continues until all information bits of the PT is conveyed successfully to the PR. When this transmission is completed, the channel is idle and the SU can solely use it until the next time the PT starts transmitting, and the same process repeats. Note that the SU refrains from cooperating if the PU performance suffers as a result. Clearly, in the second slot, the power ratio between the superpositioned PT and ST data is an important factor governing the performance of both the PU and SU. When it favors the PT data, the ST data rate suffers. However, this also leads to higher data rate for the PT and hence a shorter PT transmission duration and longer channel idle time for the ST to transmit. To study the effect of this power ratio is one of our goals in Chapter 3.

To best of our knowledge, our work in Chapter 3 was the first to investigate such a problem. Nevertheless, maximizing the performance of the SU while maintaining that of the PU had been previously studied in various literature. For example, [37] proposes a distributed power/channel allocation algorithm to maximize the down link coverage and throughput while maintaining the interference to the PU below a certain threshold. In this work, available channels are divided into two parts: one for the ST and another one for the PU. In addition, the ST does not cooperate in transmitting the PU data. In [38], adaptive user cooperation in heterogeneous cognitive relay system is considered. To maximize the throughput of the SU while maintaining that of the PU, one best relay from a pool of relays is selected, and optimal power allocation and beamforming is performed between the source and the relay, assuming that colliding packet can always be decoded. In [39], a CR model is investigated in which there are one single PT and multiple STs. In the adopted cooperative scheme, the time is divided into three unequal partitions. In
the first time slot, the PT broadcasts its message to STs only, and all STs are required to decode it. In the second time slot, the PT and all STs transmit this message simultaneously to the PR. In the third time slot, the channel is relinquished to the SUs and each SU utilizes it based on a pricing scheme. Unlike our model, the colliding packets are assumed to be decoded, and only DF without SIC is considered. Moreover, the ST experiences an inevitable delay until the third time slot. In our model, the ST can simultaneously conveys its own message to the SR, and therefore, the delay is relatively negligible. Also, since the duration of the first and the second time slots are not necessarily the same, different coding rates are required.

2.2 Device-to-Device Assisted Virtual MIMO

Utilizing a cooperating device has been shown to improve the spectral efficiency and diversity while alleviating the outage behaviour [40]. This cooperation is performed by a second wireless device (the relay) relaying the message of a first wireless device (the source). In the seminal work of Laneman et al. [40], the diversity gain achieved by this cooperation is investigated, and it is shown that full diversity can be achieved for the repetition-based DF scheme. These results have inspired numerous works on selection algorithms in the cooperative diversity area.

There is a rich body of literature on how to select the proper relay(s) to cooperate with the source(s) that can be categorized into three trends:

1. In the earlier works, e.g., [43, 46], selection of a single best relay (for a single source) based on different performance metrics including the best instantaneous channel condition [44] or the trade off between the AF error probability and the power consumption [46] is studied. For instance, in [43], one DF relay is selected when the direct link between the source and destination has low quality.

2. Later works have generalized the single relay selection concept to multiple relay selection in an effort to find optimal relay selection methods under various assumptions [47].
Most of these works consider AF relaying (e.g., [47–51]). The DF two-hop relaying technique is considered in [53–55]. To mitigate the interference, in [53], only the source and relays that reside in the same cooperation region are allowed to be grouped together, and the effect of the cooperation region size on the network sum-rate is investigated. Both [54] and [55] investigate algorithms to select \( m \) best relays out of \( N \) uniformly distributed available relays for a single source, and obtain approximations on the cumulative distribution function (CDF) of the source spectral efficiency when \( m \) relays are selected. Channel gains considered in [54] are modelled by only Rayleigh fading whereas [55] assumes that mean channel gains are known.

3. In recent works, relay selection for multiple sources is considered in [56–60]. AF relaying is considered in [56] where \( L \) relays are selected. DF selection is considered in [57–60]. In [57] and [58], orthogonal channels are assumed for each source, thus eliminating the effect of the leakage interference. In [57] and [59] a fixed number of relays (one) is selected for each source. The relay selection problem is approached from a pricing based game model in [60, 61]. In [60] each relay demands a price (virtual currency) for cooperation, and each source bids a price to recruit relays, and in addition, a competitive price adjustment process is discussed.

Although selection is not a new concept, a scarcity of studies on clustering (grouping) algorithms from a MIMO perspective is apparent. For instance, most previous works including [56–60] consider only single antenna at the receiver and fail to examine the effect of the precoding.

In a related line of work, clustering algorithms have long been studied in the context of wireless sensor networks (WSNs), e.g., the pioneering works [62–64]. These studies are mostly concerned with determining the best cluster head such that firstly, the power consumption is minimized, and secondly, the multi-hop link latency from a sensor to the cluster head is bounded. Given a predetermined clustering, [65] proposes two algorithms (based on minimum spanning tree and singular value decomposition) to select the best channels between two
clusters of sensors in a multi-hop VMIMO setting. Nevertheless, the clustering algorithms in WSNs are often aimed at maximizing the network lifetime whereas in the cellular communication systems, the diversity is exploited to maximize the spectral efficiency, and this difference leads to substantially incomparable algorithms.

Recently, the problem of clustering algorithms for VMIMO has gained increasing attention since VMIMO is recognized as a promising future trend of communication systems [9, 66, 67]. In particular, there has been a considerable number of studies on VMIMO relevant to the LTE standard both from a macrocell perspective [66, 68, 69] as well as a femtocell perspective [67]. The problem of joint grouping and precoding for fixed size VMIMO where precoding weights are continuous is studied in [70]. The uplink pair selection problem (e.g., [71–74]) is extended to multi user uplink grouping for a single destination in [75]. In this work, a fixed number of users are selected (and grouped together) such that a proportional fairness utility is maximized. Nonetheless, the cooperation between nodes is not considered, and each node transmits its individual message.

2.3 Iterative Beamforming and Clustering for Virtual MIMO

The seminal work of Cadambe and Jafar [76] proposes a profound beamforming technique termed as interference alignment and shows that the asymptotic sum-capacity of a $K$-user interference channel can be scaled as $K/2$. In this technique, all the users transmit on the co-channel while trying to map their interference footprint on a dimension orthogonal to that of the desired signal. One of the practical challenges to achieve this linearly scaled degrees of freedom, however, is that the proposed interference alignment technique requires a channel expansion (either in time or frequency or the number of receive and transmit antennas) that grows exponentially with the number of users [77]. On the contrary, the prevalent method to manage the interference (widely used in current standards including LTE and IEEE 802.11) is to transmit signals on orthogonal channels either in time, frequency, or space. The drawback of
this orthogonalization is that the asymptotic sum-rate remains constant as the number of users increases.

As a pioneering work in the area of interference alignment, Gomadam et al. \cite{78} proposed an iterative distributed algorithm based on interference alignment tailored for the high SINR. In this algorithm, the channel reciprocity is utilized to jointly design the beamforming coefficients. The interference in the reciprocal channel is decomposed and the beamforming vectors are chosen as the smallest eigen modes of this reciprocal interference. In addition, they proposed the acclaimed iterative max-SINR algorithm that outperforms this interference alignment algorithm in a low SINR regime. To this date, however, there is no general convergence proof for the max-SINR algorithm although it is shown to be convergent in some special cases \cite{79}.

Inspired by the interference alignment technique, many distributed beamforming algorithms have been developed for the MIMO interference channels without requiring the channel expansion \cite{80-88}. A rather comprehensive comparison of these algorithms is given in \cite{89}. Almost all of these methods rely on iterative solutions to find beamforming weights. These works can be divided into the following categories:

1. **Selfish Updates**: In this strategy, each user selfishly maximizes its own SINR (similar to Chapter 5).

2. **Min-Leakage \cite{76}**: This method only focused on minimizing the interference to other users. Ideally, the beamforming weights should be chosen such that the interference is zero which is not always possible.

3. **max-SINR \cite{78}**: This method is motivated by the uplink duality for the multiple access and broadcast channels. In the max-SINR algorithm, the beamforming vectors in the forward channels are the minimum mean squared error (MMSE) decoders in the reciprocal channels.

4. **MMSE \cite{82,86}**: In a separated line of work from that of the interference alignment, the mean squared error (MSE) based transceiver design for the MIMO interference channel
is proposed in \cite{82, 86}. In these works, a series of beamforming update for MWSMSE is obtained. Nonetheless, the updates for the MSE based algorithms are shown to be a generalized version of those for the max-SINR algorithm although MSE based designs use a different optimization approach \cite{89}.

5. **Adaptively Weighted MMSE \cite{83}**: This approach is very similar to the MMSE approach except that MSE terms are weighted, and these weights are adaptively determined to maximize a network sum-utility including sum-rate and proportional fairness.\footnote{The distributed version of the weighted MMSE is studied in Section 6.2} Note that the MSE based method is a special case of the weighted MMSE method.

In this thesis, the distributed beamforming methods are developed (in Chapter 6) based on the Weighted MMSE. However, as will be shown later, the current solutions in the literature are not applicable to our VMIMO scenario due the global CSI knowledge requirement. Therefore, we proposed a new set of beamforming updates that addresses this issue. To the best of the author’s knowledge, this is the first attempt to address the beamforming problem for VMIMO, and therefore, there is a lack of prior works to compare with. Therefore, as the baseline for the comparison, the celebrated max-SINR is chosen. It should be, however, noted that the max-SINR still needs that global CSI, and given this global CSI, it is expected to outperform our methods.

Although the above mentioned MIMO technique have been perceived as a promising approach to enhance the spectral efficiency of wireless systems, in practice, commonly for low cost legacy UEs only a single transmit path is implemented to preserve battery time. However, multiple single-antenna UEs may be clustered together to create a VMIMO system. There are two new major challenges that arise in this realm: firstly, since the precoding updates in the literature requires the CSIs for all the antennas on the device, they can not be applied to the scenario where UEs virtually form a multi antenna device unless a central processor with all the CSI calculates the precoding weights for each set of UEs in a VD which is highly impractical;
secondly, given a large number of UEs, an optimal subset of them must be selected to form a VMIMO device (also known as the clustering or user grouping problem).

Surveying the literature shows that there are two prevalent approaches to the UE clustering problem each with its own drawbacks when applied to a wireless system. A first approach is to employ a clustering method from the wealth of algorithms in the machine learning area (see [90] for a comprehensive review of clustering algorithms). These conventional approaches can be categorized into four classes known as partitional clustering, hierarchical clustering, individual clustering, and affinity propagation based clustering [91–94]. Almost all these approaches are explicitly or implicitly dependent to the definition of a PM [90]. Once a PM is determined, clustering could be construed as an optimization problem with a specific objective function. However, the parameters governing a wireless system (including rates) are functions of the clustering itself, and therefore, the PMs tend to change as the clusters evolve. To address this problem, a game theoretic approach can be adopted. In this method, UEs are considered as players who participate in an auction by iteratively bidding towards the clustering with the desired property (maximum sum-rate or revenue) [58–60]. However, the problem with the latter approach is that the number of the bidding rounds usually scales with the number of source UEs. Therefore, such an approach may not be practical in a scenario with large number of users.

The choice of the PM can determine the trade-off between the performance gain and the processing complexity as well as both the amount of CSI signalling and UE data to be exchanged between the cooperating nodes. Most previous work considers heuristic PMs that suits the specific problem at hand. Examples include average channel gains [95] in the coordinated multipoint (CoMP) scenario, channel gains between the APs in the overlaid femtocell scenario [96], the number of common channels shared between pairs of secondary users in the context of the cognitive radio [97], and the mutual inter-cluster interference between pairs of base stations in the CoMP scenario [98]. This observation shows that there is no ubiquitous consensus over the choice of the PM, and it is often determined by the nature of the problem.
Chapter 3

Half-Duplex Cognitive Radio with Collision Avoidance

In this chapter, a CoCR system is studied where a SU may relay data for a PU. Under the HD assumption and packet collision avoidance model, the instantaneous data rates for the PU and SU are derived as a function of the SU’s power allocation factor. We study the optimal power allocation to maximize either the SU’s instantaneous data rate or the long-term throughput contingent on not degrading the PU’s spectral efficiency. Our numerical results suggest that these two performance criteria entail drastically different optimal behaviour by the secondary user. Furthermore, we observe that employing SIC in relaying can significantly increase the instantaneous data rate, but after optimizing the power allocation factor, the simple DF strategy almost always achieves the same throughput as that of more complicated strategies.

This chapter is organized as follows. The system model is presented in Sec. 3.1. Sec. 3.2.1 analyzes the DF strategies with and without SIC. Sec. 3.2.2 analyzes the AF strategies with and without SIC. Sec. 3.4.1 illustrates a practical case in which the ST is moving and demonstrates the effect of location (channel gains) on the data rate of the SU. Sec. 3.3 discusses the throughput of the system for different cooperative strategies. Finally, Sec. 3.5 concludes this

---

1Contents of this chapter have been mainly published in [10] (Digital Object Identifier: 10.1109/INF-COMW.2011.5928880).
3.1 System Model and Problem Statement

We consider a CoCR system as depicted in Fig. 3.1. In this model, there is a PU with one transmitter (PT) and one receiver (PR). The PT has a message to transmit, and the PR receives and decodes that message. There is also a SU with one transmitter (ST) and one receiver (SR). For convenient indexing, PT, PR, ST, SR are denoted by nodes 1, · · ·, 4.

Unlike previous studies [3, 16–18, 23], the ST is assumed to be a relay which can only operate in half-duplex mode (similar to the model adopted in [32]). The PR and the SR cannot decode colliding packets from different sources; when a collision occurs the receiver drops both packets. If the ST cooperates, then the time is divided into alternating slots. In odd slots, the PT transmits its own signal whereas in even slots, the ST transmits a superimposed signal containing its own message and a version of what it received from the PT in the previous slot. We also define the power allocation factor $\beta$ as the portion of the ST’s transmission power that
is dedicated to transmit PT’s signal.

To mathematically model this communication channel, let $\tilde{X}_i$ indicate the transmitted signal by node $i$, and $\tilde{Y}_i$ indicate the received signal by node $i$, $i = 1, 2, 3, 4$. Also, let $\tilde{h}_{ij}$ be the channel gain between nodes $i$ and $j$. The relationship between transmitted signals and received signals can be expressed as

$$
\begin{bmatrix}
\tilde{Y}_2 \\
\tilde{Y}_3 \\
\tilde{Y}_4
\end{bmatrix}
= 
\begin{bmatrix}
\tilde{h}_{21} & 0 \\
\tilde{h}_{31} & \tilde{h}_{32} \\
\tilde{h}_{41} & \tilde{h}_{42}
\end{bmatrix}
\begin{bmatrix}
\tilde{X}_1 \\
\tilde{X}_2
\end{bmatrix}
+ 
\begin{bmatrix}
\tilde{Z}_2 \\
\tilde{Z}_3 \\
\tilde{Z}_4
\end{bmatrix}
$$

(3.1)

where $\tilde{Z}_i$ is an additive white Gaussian noise at node $i$ with power $E[\tilde{Z}_i^2] = \tilde{N}_i$. The channel gains $\tilde{h}_{ij}$ are assumed to be real; the extension to complex channel gains is straightforward and is ignored here. The transmitters are subjected to the power constraint

$$E[\tilde{X}_i^2] \leq \tilde{P}_i.$$

We term this channel as a Gaussian interference channel hereafter. In this model, it is assumed that the PT can communicate with the PR with the Shannon rate $R_1 = I(\tilde{X}_1; \tilde{Y}_3)$. In our model, $R_1 = 0.5 \log(1 + \text{SINR})$ where the logarithms are taken in base 2, and the SINR is the ratio of the received signal power at the PR over the total noise and interference power, i.e.,

$$\text{SINR} = \frac{\tilde{h}_{31}^2 \tilde{P}_1}{(\tilde{h}_{32}^2 \tilde{P}_2 + \tilde{N}_1)}.$$

To make the analysis more tractable, we now present the notion of the standard channel. It has been shown (see [99] for example) that by applying a scaling transformation on the parameters and input signals, the resulting changes in output signals can be compensated by constant scaling factors. That is, two Gaussian interference channels related by this scaling transformation are equivalent in terms of performance metrics.
Let (3.1) represent the original channel input output relationship. The channel characterized by the following input output relationship has the same performance metrics as that in (3.1).

\[
\begin{bmatrix}
Y_2 \\
Y_3 \\
Y_4
\end{bmatrix} = \begin{bmatrix}
a & 0 \\
1 & c \\
b & 1
\end{bmatrix} \begin{bmatrix}
X_1 \\
X_2
\end{bmatrix} + \begin{bmatrix}
Z_2 \\
Z_3 \\
Z_4
\end{bmatrix},
\]

(3.2)

where

\[
a = \tilde{h}_{21} \sqrt{\frac{N_3}{N_2}}, \quad b = \frac{\tilde{h}_{41} \sqrt{N_3}}{\tilde{h}_{31} \sqrt{N_4}}, \quad c = \frac{\tilde{h}_{32} \sqrt{N_4}}{\tilde{h}_{42} \sqrt{N_3}},
\]

and the new transmitted signals are

\[
X_1 = \frac{\tilde{h}_{31} \tilde{X}_1}{\sqrt{N_3}}, \quad X_2 = \frac{\tilde{h}_{42} \tilde{X}_2}{\sqrt{N_4}}.
\]

As a result of this transform, the new transmitter powers will be

\[
P_1 = \frac{\tilde{h}_{31}^2 \tilde{P}_1}{N_3}, \quad P_2 = \frac{\tilde{h}_{42}^2 \tilde{P}_2}{N_4},
\]

and the noise powers are normalized to one. For convenience, we will analyze the standard channel hereafter; nonetheless, the results are easily transformable to the form that contains the original channel’s parameters.

Let \( R_1 \) indicate the information theoretic rate for the PU. Similarly, let \( R_2 \) be the SU’s rate by which the ST conveys the information to the SR. As mentioned before, the PU has the right to use the spectrum, and its rate may not be degraded by the interference incurred by the SU. The data rate of the PU when it fully utilizes the channel is

\[
R_1^* = \frac{1}{2} \log(1 + P_1).
\]

(3.3)
If the ST transmits concurrently with power $P_2$, the rate of the PU diminishes to

$$R_1 = \frac{1}{2} \log \left( 1 + \frac{P_1}{1 + e^2 P_2} \right).$$

The SU is, however, a smart cognitive user that can cooperate with the PU. It can play a twofold role: On the one hand, if the ST selflessly allocates all of its power to relay the message of the PT, it can enhance the performance of the PU while achieving zero rate for the ST; on the other hand, if the ST selfishly allocates all of its power to transmit its message, the rate of the PU decreases.

Initially, our objective is to design schemes for PUs and SUs such that the transmission rate of the PU remains as that in (3.3), as if there were no interference by the SU while maximizing the rate of the SU. Later (Section 3.3), we will also consider allowing the PU rate to be increased (decreased), allowing more (less) channel idle time for the SU.

### 3.2 Achievable Rates of the Cooperative CR

In this section, we investigate the achievable data rates of the CoCR modeled in Section 3.1 under DF and AF schemes.

#### 3.2.1 Decode-and-Forward Strategies

In the absence of the SU, let $2n$ be the number of channel uses for a packet of data conveying $b$ bits of uncoded information. When the SU cooperates, the PU only transmits half of the time, i.e., in $n$ channel uses, and the SU transmits in the second half of the time. Each half time is called one time slot. It is noteworthy that $n$ is assumed to be large enough such that the Shannon capacity results can be applied here. In more precise words, in first time slot, PT
transmits

\[ x^n_1 = (x_{1,1}, x_{1,2}, \cdots, x_{1,n}) \]  

(3.4)

where \( x^n_1 \) denotes the PT's message (packet) vector with length \( n \), and \( x_{1,i} \) represents the message of the PT at time instance \( i \). In the second time slot, the ST transmits

\[ x^n_2 = (x_{2,n+1}, x_{2,n+2}, \cdots, x_{2,2n}). \]  

(3.5)

When it is clear from the context we drop the time index \( i \) for \( x_{1,i} \) or \( x_{2,i} \). For a given time instance \( i \in \{1, \cdots, n\} \), let \( X'_i \) be an i.i.d. normal random variable with mean zero and unit variance. The output of the PT can be represented as

\[ x_1 = \sqrt{P_1} x'_1. \]  

(3.6)

In the DF scheme, the ST decodes the message of the PT transmitted in the first time slot, i.e., \( x^n_1 \). In the second time slot, at time index \( j = i + n \), the transmitted symbol by the ST can be expressed as

\[ x_2 = \sqrt{\beta P_2} x'_1 + \sqrt{\bar{\beta} P_2} x'_2 \]  

(3.7)

where \( x'_2 \) is a Gaussian i.i.d. random variable with mean zero and unit variance. As can be seen, the ST’s power is split into two parts: first part, \( \beta P_2 \), is used to forward the message of the PT; the second part, \( \bar{\beta} P_2 \), is used to transmit the message of the ST where \( \bar{\beta} = 1 - \beta \). The power allocation factor \( \beta \) plays a rate tuning role: intuitively, increasing \( \beta \) (allocating more power to the PT’s message and less power to the ST’s message) results in increasing \( R_1 \) and decreasing \( R_2 \).

Let \( y^{(1)}_k \) be the received signal at node \( k \) in the first time slot at an arbitrary time index \( i \),
and \(Y_{k}^{(2)}\) be the received signal at node \(k\) in the second time slot at the respective time index \(i + n\). Therefore, the received signals at the PR and the SR can be expressed as

\[
Y_3 = \begin{bmatrix} Y_3^{(1)} \\ Y_3^{(2)} \end{bmatrix} = \begin{bmatrix} \sqrt{P_1} \\ c\sqrt{\beta P_2} \end{bmatrix} X_1' + \begin{bmatrix} 0 \\ c\sqrt{\beta P_2} \end{bmatrix} X_2' + \begin{bmatrix} Z_3^{(1)} \\ Z_3^{(2)} \end{bmatrix}, \tag{3.8}
\]

\[
Y_4 = \begin{bmatrix} Y_4^{(1)} \\ Y_4^{(2)} \end{bmatrix} = \begin{bmatrix} b\sqrt{P_1} \\ \sqrt{\beta P_2} \end{bmatrix} X_1' + \begin{bmatrix} 0 \\ \sqrt{\beta P_2} \end{bmatrix} X_2' + \begin{bmatrix} Z_4^{(1)} \\ Z_4^{(2)} \end{bmatrix}, \tag{3.9}
\]

where \(Z_k^{(i)}\) is the white Gaussian noise at node \(k\) in time slot \(i\) with unit variance.

Since we impose the condition that the message of the PT must be decoded at the ST, the transmission rate of the PT should be less than the point-to-point capacity between the PT and the ST. This capacity can be written as \(\frac{1}{2} C(a^2 P_1)\) where the \(\frac{1}{2}\) factor is due to the fact that the PT transmits half of the time, and \(C(x) = \frac{1}{2} \log(1 + x)\). In addition, under this cooperative scheme, we implicitly assume that the rate of the PT is as described in (3.3); the following constraint must be satisfied:

\[
\frac{1}{2} \times C(a^2 P_1) > R^\ast_1 \Rightarrow a^2 > 2 + P_1. \tag{3.10}
\]

In the rest of this chapter we investigate three methods of decoding: first, decoding without interference cancellation; second, decoding with SIC at the PR; third, decoding with SIC at the SR. Also, let the rate \(R_1\) be associated to the number of bits per channel use that \(X_1^m\) conveys, and similarly, the rate \(R_2\) be associated to that of \(X_2^m\).

**Decode-and-Forward with no SIC**

In the DF method with no SIC each receiver only decodes its intended message while treating the message of the other transmitter as interference. The following theorem characterizes an achievable rate region under this scheme.
Theorem 1 In the channel described in Section 3.1, a reliable communication is feasible if

\[
R_1 \leq \frac{1}{4} \log \left(1 + P_1 + c^2 P_2 + c^2 \frac{\bar{\beta} P_1 P_2}{1 + c^2 \beta P_2}\right), \tag{3.11}
\]

\[
R_2 \leq \frac{1}{4} \log \left(1 + \frac{\bar{\beta} P_2}{1 + \beta P_1}\right), \tag{3.12}
\]

where \(0 \leq \beta \leq 1\).

\textbf{Proof:} See Appendix B.1. \qed

\section*{Decode-and-Forward with SIC at the PR}

In the second time slot, the PT is broadcasting two messages, \(X'_1\) and \(X'_2\), to the PR and the SR. When the cross channel gain \(c\) is greater than one, the channel formed between the ST, the PR, and the SR is a degraded broadcast channel; that is the signal received by the SR is a degraded version of that received by the PR. In this case, let the term weaker receiver be used for the SR and the term stronger receiver be used for the PR. Cover [100] and Gallager [101] showed that the capacity achieving strategy for the degraded broadcast channel is as follows: The transmitter superimposes the message of the weaker receiver onto the message of the stronger receiver. The stronger receiver first decodes the message of the weaker receiver, and then utilizes this knowledge to mitigate the interference when it decodes its own message. The weaker receiver only decodes its intended message. When the channel is Gaussian, the superposition coding is simply performed by linearly combining two codewords.

Having this guide line in mind, the following scheme utilizes the advantage of successive cancellation at the PR: let the PR first decode the message of the ST \((X'_2)\) and then subtract it from the received signal \(Y_3\). In other words, the PR decodes \(X'_2\) first and then \(X'_1\), and the SR only decodes \(X'_2\).
Theorem 2  In the channel described in Section 3.1, a reliable communication is possible if

\[ R_1 \leq \frac{1}{4} \log(1 + P_1 + c^2 \beta P_2), \quad (3.13) \]
\[ R_2 \leq \min \left\{ \frac{1}{4} \log \left( \frac{1 + P_2}{1 + \beta P_2} \right), \right. \]
\[ \left. \frac{1}{4} \log \left( \frac{1 + P_1 + c^2 P_2 + c^2 \beta P_1 P_2}{1 + P_1 + c^2 \beta P_2} \right) \} \quad (3.15) \]

for \( 0 \leq \beta \leq 1. \)

Proof: See Appendix B.2

Decode-and-Forward with SIC at the SR

The DF method with SIC at the SR is analogous to that with SIC at the PR (Section 3.2.1) except that the interference cancellation is performed at the SR instead of the PR. The following theorem states the rates achieved by this method.

Theorem 3  For the channel model described in Section 3.1, a reliable communication is feasible if

\[ R_1 \leq \min \left\{ \frac{1}{4} \log \left( \frac{1 + P_1 + c^2 P_2 + c^2 \beta P_1 P_2}{1 + c^2 \beta P_2} \right), \right. \]
\[ \left. \frac{1}{4} \log \left( \frac{1 + b^2 P_1 + P_2 + b^2 \beta P_1 P_2}{1 + \beta P_2} \right) \} \quad (3.16) \]
\[ R_2 \leq \frac{1}{4} \log(1 + \beta P_2) \quad (3.18) \]

for \( 0 \leq \beta \leq 1. \)

Proof: See Appendix B.3
3.2.2 Amplify-and-Forward Strategies

Similar to the DF scheme, in AF scheme, the PT transmits its message in the first time slot. Let

\[ X_1 = \sqrt{P_1} X'_1. \]  

(3.19)

be the message transmitted by the PT at time index \( i \) where \( X'_1 \) an i.i.d. Gaussian random variable with zero mean and unit variance. The received signal by the ST in this time slot is

\[ Y_2 = Y_2^{(1)} = a \sqrt{P_1} X'_1 + Z_2^{(1)} \]  

(3.20)

where \( Z_2^{(1)} \) is the white Gaussian noise with unit variance.

In the second time slot the PT remains silent while the ST first scales the received signal \( Y_2 \) to normalize its power and then, allocates \( \beta \) portion of its power to forward the scaled version of \( Y_2 \). In more precise words, in the second time slot at time index \( j = i + n \), the ST transmits

\[ X_2 = \sqrt{\beta P_2} \frac{Y_2}{\|Y_2\|} + \sqrt{\bar{\beta} P_2} X'_2. \]  

(3.21)

where \( X'_2 \) is a Gaussian i.i.d. random variable with mean zero and unit variance, and \( \|Y_2\| = \sqrt{E[|Y_2|^2]} = \sqrt{1 + a^2 P_1} \). As can be seen, the power of the ST is split into two parts: first part, \( \beta P_2 \), is used to forward the received signal from the PT, the second part \( \bar{\beta} P_2 \) is used to transmit the message of the ST. Similar to the DF case, the power allocation factors \( \beta \) in the interval \([0, 1]\) and \( \bar{\beta} = 1 - \beta \) are to be determined later.

Similar to (3.8) and (3.9), the received signals at the PR, \( Y_3 \), and the SR, \( Y_4 \), can be expressed in matrix forms.

Amplify-and-Forward with no SIC

Similar to the DF scheme with no SIC if the PR and the SR solely decode their respective messages, i.e., \( X'_1 \) and \( X'_2 \). The following theorem characterizes the data rates achievable
under this scheme:

**Theorem 4**  *For the channel model described in Section 3.1, a reliable communication is feasible if*

\[
R_1 \leq \frac{1}{4} \log \left( \frac{(1+c^2P_2)((1+a^2+c^2P_1)P_1+1)-\beta a^2c^2P_2^2P_2}{1+a^2+c^2P_2+a^2c^2P_1P_2-\beta a^2c^2P_1P_2} \right), \\
R_2 \leq \frac{1}{4} \log \left( \frac{1+P_2}{1+\beta P_2} \right)
\]  
(3.22) (3.23)

*for* \(0 \leq \beta \leq 1\).

A similar proof method to that of the Theorem 1 is imitated here, and so, we omit the proof.

**Amplify-and-Forward with SIC at the PR**

Similar to Section 3.2.1 in AF method with SIC at the PR, the PR first decodes the message of the ST and then, the PR subtracts it from the overall received signal to obtain a more clear version. The SR, on the other hand, only decodes the intended signal \(X_2'\). The following theorem establishes the achievable rate region under this scheme.

**Theorem 5**  *A reliable communication in communication scenario modeled in Section 3.1 is feasible if*

\[
R_1 \leq \frac{1}{4} \log \left( \frac{1+P_1+\alpha^2P_1+\alpha^2P_2^2+\beta c^2P_2(1+P_1+\alpha^2P_1)}{1+\alpha^2P_2+\beta c^2P_2} \right), \\
R_2 = \min \left\{ \frac{1}{4} \log \left( \frac{1+(1+\alpha^2)(P_1+c^2P_1)P_2+c^2P_2+\alpha^2P_2^2+\alpha^2P_2^2P_2+\beta c^2P_2(1+\alpha^2P_1+P_1)}{1+(1+\alpha^2)P_1+\alpha^2P_2^2+\beta c^2P_2(1+\alpha^2P_1+P_1)} \right), \right. \\
\frac{1}{4} \log \left( \frac{1+\alpha^2P_1+P_2+\alpha^2P_1P_2}{(1+\alpha^2P_1)(1+\beta P_2)} \right) \left. \right\}
\]

*for* \(0 \leq \beta \leq 1\).

The proof follows the same steps as those of the Theorem 2 and therefore, it is omitted here.
Amplify-and-Forward with SIC at the SR

Similar to the method in Section 3.2.1, in the AF method with SIC at the SR, the SR first decodes the signal of the PT $X_1^n$, and then subtracts it from the received signal to reduce the SINR. The PR on the other hand, only decodes the intended message $X_1^n$. The following theorem provides an achievable rate region under this scheme.

**Theorem 6** In the communication channel described in Section 3.1 an error-free communication is possible if

\[
R_1 \leq \min \left\{ \frac{1}{4} \log \left( \frac{1 + P_1 (a^2 + b^2)(1 + P_2) + a^2 b^2 P_1^2 + P_2 + a^2 b^2 P_2^2 P_1}{1 + a^2 P_1 + a^2 b^2 P_1 P_2 \beta} \right) \right\}, \tag{3.24}
\]

\[
\frac{1}{4} \log \left( \frac{(1 + c^2 P_2)(1 + a^2 + a^2 P_1) P_2 + 1 - \beta a^2 c^2 P_2^2 P_1}{1 + a^2 P_1 + a^2 c^2 P_1 P_2 - \beta a^2 c^2 P_1 P_2} \right) \right\}, \tag{3.25}
\]

\[
R_2 \leq \frac{1}{4} \log \left( 1 + \frac{\beta P_2 (1 + a^2 P_1)}{1 + a^2 P_1 + \beta P_2} \right) \tag{3.26}
\]

for $0 \leq \beta \leq 1$.

The proof of this theorem follows same steps as those of the Theorem 3, and we omit it here.

To conclude this section, we define a special power allocation factor $\beta_R$ that maximizes the rate and compute it for different strategies presented in this section.

**Definition 3.2.1** A rate maximizing power allocation factor $\beta_R$ is the power allocation factor $\beta$ adopted by the ST that maximizes the rate achieved by the SU $R_2$, while maintaining the interference-free rate for the PU as described in (3.3).

It can be verified that the rates $R_1$ obtained in Theorems 1 to 6 are monotonically increasing functions of $\beta$ whereas the ST’s rate, $R_2$, obtained in those theorems are monotonically decreasing functions of $\beta$. Moreover, the data rate of the PT must at least be equal to that in (3.3), and the optimal power allocation can be obtained by comparing (3.3) and $R_1$ in each
Table 3.1: List of cooperative strategies studied in this chapter.

<table>
<thead>
<tr>
<th>Legend</th>
<th>Cooperative Strategy</th>
<th>Rate Maximizing Power Allocation Factor $\beta_R$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(i)</td>
<td>DF with no SIC</td>
<td>$\frac{P_1 + P_2 + c^2 P_1 P_2 + c^2 P_2 P_1}{c^2 P_2 + c^2 P_1 P_2 + c^2 P_2 P_1}$</td>
</tr>
<tr>
<td>(ii)</td>
<td>DF with SIC at the PR</td>
<td>$\frac{P_1 + P_2}{(P_1^2 + P_2^2)}$</td>
</tr>
<tr>
<td>(iii)</td>
<td>DF with SIC at the SR</td>
<td>$\min \left{ \frac{1}{1+(1+a^2)(P_1+c^2 P_1 P_2)+c^2 P_2+a^2 P_1^2+c^2 P_2^2 P_1} \right}$</td>
</tr>
<tr>
<td>(iv)</td>
<td>AF with no SIC</td>
<td>$\frac{(1+(1+a^2)P_1+a^2 P_2^2)/(c^2 P_2(a^2-1-P_1))}{1+(1+a^2)(P_1+c^2 P_1 P_2)+c^2 P_2+a^2 P_1^2+c^2 P_2^2 P_1} \right}$</td>
</tr>
<tr>
<td>(v)</td>
<td>AF with SIC at the PR</td>
<td>$\min \left{ \frac{1}{1+(a^2)(P_1+c^2 P_1 P_2)+c^2 P_2+a^2 P_1^2+c^2 P_2^2 P_1} \right}$</td>
</tr>
<tr>
<td>(vi)</td>
<td>AF with SIC at the SR</td>
<td>$\min \left{ \frac{1}{1+(1+a^2)(P_1+c^2 P_1 P_2)+c^2 P_2+a^2 P_1^2+c^2 P_2^2 P_1} \right}$</td>
</tr>
</tbody>
</table>

scheme. These optimal power allocations $\beta_R$ are computed and listed in Table 3.1. Note that these values for $\beta_R$ are only valid if they are in the interval $[0, 1]$; otherwise, the ST refrains from transmitting, and the PT utilizes both time slots.

### 3.3 Throughput Analysis of the Cooperative CR

In the previous section, we studied the data rate of the SU and showed how it varies with respect to different channel parameters when the SU is cooperating with the PU. An equally important performance metric for the wireless channels is the throughput or the average spectral efficiency (ASE), here defined as the time averaged number of uncoded bits transmitted over a relatively long period of time. As illustrated in Fig. 3.2, let the time between two consecutive messages of the PU be $S$. We assume the required throughput of the PU is $\zeta S R_1^*$ bits where $0 < \zeta < 1$ is the channel utilization factor (duty cycle), and $R_1^*$ is the maximum point to point data rate obtained in (3.3), i.e., the PT can transmit the message to the PR in duration $\zeta S$. Hence, for the rest of the time, $(1 - \zeta)S$, the channel is idle and can be utilized by other users.

Consider a slotted transmission as modeled in Section 3.1. If the ST cooperates with the PT, the rate at which the PU can convey the data changes, and so does the time the PT requires to transmit $\zeta S R_1^*$ bits. Let $R_1(\beta)$ be the rate achievable by the PU when the ST’s power allocation
factor is $\beta$ (see (3.7), (3.21)). Also, let $R_2(\beta)$ be the rate of the SU with power allocation factor $\beta$. Note that $R_2(0)$ is a special case meaning that the ST is either transmitting with full rate $0.5 \log(1 + P_2)$ if the channel is ideal or not transmitting if the PT is transmitting.

Denote by $T$ the time duration that the PU requires within $S$, to achieve the same throughput $\zeta S R_1^*$ with the rate $R_1(\beta)$. Then, the following constraint must hold to maintain the throughput of the PU intact:

$$T R_1(\beta) = \zeta S R_1^*.$$  \hspace{1cm} (3.27)

Depending on the chosen $\beta$, $T$ may be less or greater than $\zeta S$. In other words, the SU can adjust the PU’s total transmission time. Note that since $T < S$, $R_1(\beta)$ must be greater than $\zeta R_1^*$.

The overall throughput of the SU when it cooperates with the PU can be computed as
follows:

$$\gamma_2 = \frac{1}{S} (TR_2(\beta) + (S - T)R_2(0))$$

$$= \frac{\zeta R_1}{R_1(\beta)} (R_2(\beta) - R_2(0)) + R_2(0).$$  \hspace{1cm} (3.28)

By substituting the $R_1(\beta)$ and $R_2(\beta)$ expressions from Theorems 1 to 6 into (3.28), a closed-form expression for the SU throughput can be obtained.

**Definition 3.3.1** A throughput maximizing power allocation factor $\beta_{Tp}$ is the power allocation factor $\beta$ that maximizes the throughput of the SU $\gamma_2$ in (3.28), subject to maintaining the throughput of the PU as that in (3.27).

It is intractable to obtain analytical forms for $\beta_{Tp}$, since $\gamma_2$ contains logarithms and polynomials. However, the availability of a closed-form formula for the SU’s throughput allows us to compute $\beta_{Tp}$ by numerical solvers.

The power allocation factors $\beta_R$ and $\beta_{Tp}$ tend to exhibit opposing behaviors, hence changing the SU’s strategies based on the performance metric. To illustrate this phenomenon, the effect of the SU transmission power on $\beta_R$, $\beta_{Tp}$, and its throughput $\gamma_2$ is shown in Fig. 3.4. As can be seen, the power allocation factor $\beta_R$ decreases as $P_2$ increases since only a constant amount of power is needed to maintain the rate of the primary user as the point-to-point rate. Unlike $\beta_R$, the power allocation factor $\beta_{Tp}$ increases as $P_2$ increases. For a relatively small $P_2$, the SU adopts a $\beta_{Tp}$ less than $\beta_R$ which in turn increases the time $T$ since $R_1(\beta_{Tp}) < R_1(\beta_R)$. The reason for this choice is that for small $P_2$, the maximum data rate of the SU is small and the SU can alleviate its throughput if it uses all of the available time. As $P_2$ increases (e.g., $P_2 > 2.5$ in Fig. 3.4), the maximum rate of the SU increases, and the SU adopts $\beta_{Tp}$ greater than $\beta_R$, hence decreasing the time $T$ and consequently, increasing the channel idle time in which the SU can transmit with a high rate and achieve a high throughput.
3.4 Numerical Illustrations

3.4.1 Data Rate for a Moving Cooperative CR

In this section, we illustrate the derived date rates for different locations of the cognitive radio and channel parameters. We adopt a path loss model in which the channel gains are inversely proportional to the distance

\[ \tilde{h}_{ij} = d_{ij}^{r/2} \]  

(3.29)

where \( r \) is the path loss exponent. In this chapter we assume \( r = 4 \), which is a practical choice in areas with a low density of scattering objects. For the first scenario we consider a cognitive radio moving in the path depicted in Fig. 3.3(a). It can be seen that DF generally performs better than AF although there are rare cases where AF outperforms DF when the ST is located far from the PT and the decoding condition in (3.10) is not met. The reason is that in the AF methods, a portion of the power of the ST is being used to forward the receiver thermal noise at the ST, \( Z_2^{(1)} \). In DF on the other hand, the whole \( \beta \) portion of the power is efficiently used to convey the message of the PT. It is noteworthy that in some cases where the decoding of the message of the PR is not possible, i.e., \( a^2 < 2 + P_1 \), AF still can guarantee a nonzero rate for the ST.

To gain an insight on the effect of different channel parameters on the performance of the SU, we take a closer look at the data rate in DF methods. In DF with SIC at the PR it can be shown that if \( c^2 + P_1^2 > 1 \) (which is a practical condition since the normalized power \( P_1 \) is usually large) \( R_2 \) in Theorem 2 can be computed as

\[ R_2 = \frac{1}{4} \log \left( 1 + \frac{c^2 (1 + P_2)}{c^2 + P_1 + P_2^2} \right) \]

if \( \tilde{h}_{31}^2 \tilde{P}_1 \tilde{N}_3 + \tilde{h}_{31}^4 \tilde{P}_2^2 < \tilde{h}_{32}^2 \tilde{P}_2 \tilde{N}_3 \) (and \( R_2 = 0 \) otherwise). This equation suggests that in order to increase the data rate of the SU, we can either locate the ST in proximity of the PR, or we...
can increase the power of the ST or we can decrease the power of the PT.

### 3.4.2 Effect of Channel Parameters on Throughput

In this section, we illustrate the effect of different channel parameters on the throughput of the CoCR.

Fig. 3.5 shows the throughput of the SU for different values of $P_2$ for the cooperative schemes discussed in Secs. 3.2.1 and 3.2.2. As shown, for all schemes, the throughput naturally increases as $P_2$ increases. It is interesting, however, that the relatively more complicated DF scheme with SIC at the PR has no advantage over the simple DF with no SIC and both achieved the highest throughput. The second highest throughput is obtained by the AF with SIC at the PR scheme. Unlike for DF, some SIC helps AF. Noting that the lowest performance is for AF with SIC at the SR together with DF with SIC at the SR, we conclude that SIC at the SR is unhelpful.

Fig. 3.6 displays a comparison between the throughputs of different cooperative schemes, for a low $P_2$, when the cross-over channel gain $c$ is varying. As shown, the throughput increases as $c$ increases, regardless of the strategy. As predicted, DF with SIC at the PR outperforms all other methods, although simple DF performs closely to it (only 1.6% lower at $c = 6$). As before, the AF and DF with SIC at the SR perform equally inferior to all other schemes. For small $c$, the PU throughput is almost constant because the power in this case is relatively low and so is the cooperative capability of the SU, so that the ST remains silent in the whole duration $\zeta_S$ and transmits only afterward.

Fig. 3.7 demonstrates the throughput of different strategies versus the cross over channel gain $c$ when the $P_2$ is relatively high. We note that, when $c < 1$, the cross channel between the ST and the PR is weaker than the forward channel between the ST and the SR and the requirement of decoding the message of the SU at the PR is a bottleneck for the rate of the SU, thus reducing the rate and the throughput of the SU. Again, we observe that medium and large channel gain $c$ (e.g., $1 < c < 6$ in this figure), the DF with no SIC and the DF with SIC at the
both achieve the same throughput. Moreover, when $c < 1$, DF with no SIC performs the same as DF with SIC at the SR in terms of throughput. Therefore, the simple DF has the best performance of both regions of $c$.

3.5 Chapter Summary

In this chapter, we studied a four-node CoCR system in which the SU is allowed to have nonzero rate only if the performance the PU is not degraded. We considered the instantaneous data rate and average throughput (ASE) as performance criteria. Under the collision model, the optimal power allocation for the ST was obtained and the conditions under which the SU is allowed to cooperate is formulated. It was observed that more sophisticated methods with SIC at the PR or the SR tend to achieve significant higher data rate as compared to that with less sophisticated methods with no SIC. We then analyzed the ASE of the SU provided that the ASE of the PU does not degrade. As a counter intuitive observation, we observed that a simple DF strategy with no SIC achieves almost the same ASE as the more complicated strategies with SIC. This phenomenon becomes more evident when the SU transmission power increases.
Figure 3.3: (Legend corresponds to the strategies listed in Table 3.1) (a) Geographical locations of the users. In this figure, the locations of the PT, PR, and SR are marked with labeled blue circles, and also, the initial location of the ST is shown by a red circle. The path along which the ST travels is shown by the red solid line. (b) Data rates achievable for the SU when the ST travels along the path in left hand side figure. The channel gains are computed according to path loss model. Transmitter powers and noise powers are $\hat{P}_1 = 10$, $\hat{P}_2 = 60$, $\hat{N}_2 = \hat{N}_3 = \hat{N}_4 = 0.1$. 

(a) CR movement path.

(b) Secondary user's data rate.
Figure 3.4: Throughput illustration for the DF strategy with SIC at the PR for different values of $P_2$. The channel parameters are as follows: $P_1 = 3$; $a = 3$; $b = 2$; $c = 3$; the channel utilization factor $\zeta = 0.4$. $\beta_R$ is the power allocation factor that maximizes the data rate of the SU while maintaining the data rate of the PU. $\beta_{Tp}$ is the power allocation factor that maximizes the throughput of the SU while maintaining the throughput of the PU intact. $\gamma_2$ is the throughput of the SU.
Figure 3.5: *(Legend corresponds to the strategies listed in Table 3.1)* Throughput versus $P_2$ for six different strategies discussed in Sec. 3.2.1 and Sec. 3.2.2. The channel parameters are as follows: $P_1 = 3$; $a = 3$; $b = 2$; $c = 3$; the channel utilization factor $\zeta = 0.4$. In this figure, (i) lays on top of (ii), and (iii) lays on top of (vi).
Figure 3.6: (*Legend corresponds to the strategies listed in Table 3.1.*) Throughput versus the channel gain $c$ for six different strategies discussed in Sec. 3.2.1 and Sec. 3.2.2. The channel parameters are as follows: $P_1 = 3; P_2 = 2; a = 3; b = 2; c = 3; \zeta = 0.4$. In this figure, (iii) and (vi) lay on top of each other.

Figure 3.7: (*Legend corresponds to the strategies listed in Table 3.1.*) Throughput versus the channel gain $c$ for six different strategies discussed in Sec. 3.2.1 and Sec. 3.2.2. The channel parameters are as follows: $P_1 = 3; P_2 = 10; a = 3; b = 2; c = 3; \zeta = 0.4$. In this figure, (iii) and (vi) lay on top of each other.
Chapter 4

Half-Duplex Cognitive Radio

In Chapter 3 a CoCR was studied with the constraint that receivers can not decode colliding packets. In this chapter, this constraint is relaxed, and the HD CoCR is studied for which a novel cooperation method is proposed. The proposed cooperation method addresses the gene requirement issue in [33] by having the SU partially decode the message of the PU in the first transmission block and forward it in the next block. The SU also employs the DPC to mitigate the interference caused by the PU. Unlike prior works with complex coding schemes (e.g., [22, 23, 32]), in this proposed scheme, there is only one power allocation parameter for each user to tune.

When the SU fully cooperates with the PU, the CoCR model used in this chapter is reduced to a three-node relay channel. In this case, the performance of the developed scheme is compared with that of the seminal DDF scheme proposed in [1], and cases where the scheme studied in this chapter outperforms the DDF scheme are illustrated.

4.1 System Model

As illustrated in Fig. 4.1 a four-node communication scenario is considered in this work: node 1 is the PU transmitter; node 2 is the SU transmitter which is also capable of overhearing the channel; node 3 is the PU receiver; and finally, node 4 is the SU receiver. To express the channel
input and output relationship, similar to Chapter 3 the notion of the standard channel [99] is adopted for the Gaussian channel:

\[ Y_2 = aX_1 + Z_2; \]  
\[ Y_3 = X_1 + cX_2 + Z_3; \]  
\[ Y_4 = bX_1 + X_2 + Z_4 \]  

(4.0a) (4.0b) (4.0c)

where \( X_i \) and \( Y_i \) are the transmit signal and the receive signal of node \( i \), respectively, and \( Z_i \) represents the Gaussian noise at node \( i \) whose power is normalized to one, i.e., \( E[|Z_i|^2] = 1 \). It is also assumed that transmit signals \( X_i \)'s are power constrained, that is, \( E[|X_i|^2] \leq P_i \).

Each user is assumed to have a buffering queue. When a file (data) arrives at the PU's queue, the PU utilizes the channel to offload this data, and until the next file arrives, the PU remains idle. This process of utilizing the channel and then becoming idle is called a cycle, which is periodically repeated. Let the average duty cycle (ratio of the busy time to the cycle duration) of the PU in the absence of the SU be \( \xi \). This duty cycle is achieved when the PU transmits with the rate \( r_{pu}^* = 0.5 \log_2(1 + P_1) \). Therefore, the ASE of the PU is \( \xi r_{pu}^* \).

Each cycle comprises a large number of blocks. The block lengths are equal, and each block accommodates enough channel uses (symbols) so that the Shannon capacity is valid for it. When the PU is busy, the SU may cooperate with it by the scheme illustrated in Fig. 4.2. After the PU becomes idle, the SU fully utilizes the channel with rate \( 0.5 \log_2(1 + P_2) \).

### 4.2 Physical Layer Cooperation Scheme

Actions taken by each node (during the busy period of the PU) can be understood if two consecutive blocks are considered together. Without loss of generality, let the first of these two blocks be the odd block, and the second of them be the even block. As illustrated in Fig. 4.2, the encoding and decoding for the Gaussian case can be summarized as follows: The message
Figure 4.1: A diagram representation for the 4-node cooperative cognitive radio. Nodes 1 and 3 are the transmitter and the receiver for the [PU] respectively. Similarly, nodes 2 and 4 are the transmitter and the receiver for the [SU] respectively.

of the [PU] transmitter (node 1) is divided into two sets whose power-normalized codewords are represented by $c_{11}$ and $c_{12}$. The first set is a non-cooperative message set whereas the second set is the cooperative message set. The cooperative message $c_{12}$ is conveyed to the destination with the help of the [SU] transmitter (node 2). In the odd block, node 1 transmits a linear combination of these two codewords. It allocates only enough power (determined by $\bar{\alpha} = 1 - \alpha$) to the cooperative message $\sqrt{\bar{\alpha}}P_{1}c_{12}$ such that it can be decoded at node 2. In the even block, node 1 transmits only the codeword $c_{12}$. The [SU] transmitter (node 2) listens and decodes $c_{12}$ in the odd block. In the even block, it linearly combines the message of the [PU] ($c_{12}$) with the DPC [4] version of its own codeword $c_{2}$. Therefore, not only does node 2 cooperate with the node 1 (PU), also knowing the codeword of node 1 in the even block ($c_{12}$), it mitigates the interference caused by the [PU] to the [SU]. The parameter $\beta \in [0, 1]$ is the power allocation factor indicating the portion of power that the [SU] transmitter is willing to allocate to relay the message of the [PU]. The [PU] receiver (node 3) performs a successive decoding at the end of the even block. It first decodes $c_{12}$ by combining received signals from nodes 1 and 2, and then, by removing $c_{12}$ from the received signal, it decodes $c_{11}$. Finally, the [SU] receiver (node 4)
Theorem 7  In the channel described in Section 4.1, when the SU is cooperating with the PU in half-duplex mode, a reliable communication is feasible if

\[
\begin{align*}
    r_{pu} &\leq \frac{1}{4} \log_2 (1 + \alpha P_1) + \\
    &\quad \frac{1}{4} \log_2 \left( 1 + \min \left\{ \frac{a^2(1-\alpha)P_1}{1+a^2\alpha P_1}, \frac{P_1+a^2\beta P_2+2c^2\sqrt{\beta P_2}}{1+c^2(1-\beta)P_2} \right\} \right) \\
    r_{su} &\leq \frac{1}{4} \log_2 (1 + (1 - \beta)P_2).
\end{align*}
\]

subject to \(\alpha\) and \(\beta\) being in the interval \([0, 1]\).

Proof: See Appendix B.4.

Fig. 4.3 compares the achievable rates of the DDF in [1], the full-duplex DF in [2], and the half-duplex DF obtained in Theorem 7 when the SU only relays the message of the PU (\(\beta = 1\)). As can be seen, the proposed scheme tends to outperform the DDF especially when \(P_2\) is small or \(a\) is large.
Figure 4.3: Comparison between achievable rates in DDF in [1], full duplex DF (FDDF) in [2], and the half duplex DF (HDDF) in Theorem 7 for \( \beta = 1 \).
4.3 Power Allocation Algorithms

The degree of freedom for the SU is in the power allocation factor $\beta$. In this section, two classes of power allocation algorithms are proposed: selfless and greedy. In the selfless algorithm, when the PU is busy, the SU allocates all of its power to relay the message of the PU. In the greedy algorithm, on the other hand, only enough power is allocated to cooperate with the PU so that the PU can offload its data queue in each cycle. In the latter scheme, when the SU cooperates with the PU, the average duty cycle of the PU is changed to $\xi^C \in [0, 1]$.

4.3.1 Selfless Power Allocation

As mentioned before, in the selfless algorithm, the SU allocates all of its power to relay the message of the PU. By substituting $\beta = 1$ in (4.1), it can be written

$$r_{pu} \leq \frac{1}{4} \log_2 \left( 1 + P_1 + (1 - a^{-2}) \left( \sqrt{P_1} + c \sqrt{P_2} \right)^2 \right)$$

(4.3)

provided that $a > 1 + c \sqrt{P_2/P_1}$. This selfless strategy, taken by the SU, often leads to a higher rate for the PU thus enabling the PU to offload its data faster. Therefore, the channel idle time is increased and there will be more time for the SU to utilize the channel.

4.3.2 Greedy Power Allocation

Let the PU and the SU rates be described as $(r_{pu}(\beta), r_{su}(\beta))$ when the SU is cooperating with the PU. The SU allocates only enough power to cooperate with the PU such that the ASE of the PU remains greater than or equal to $\xi r_{pu}^*$. This condition can be mathematically formulated as

$$\xi^C r_{pu}(\beta) \geq \xi r_{pu}^*.$$  

(4.4)
The ASE of the SU, $\gamma_{su}$, can be obtained as

$$\gamma_{su}(\beta) = \xi C r_{su}(\beta) + (1 - \xi C) r_{su}(0)$$

$$= \frac{\xi r_{pu}^*}{r_{pu}(\beta)} (r_{su}(\beta) - r_{su}(0)) + r_{su}(0) \quad (4.5)$$

in the case of equality in (4.4) where $r_{su}(0)$ is the rate of the SU during the PU’s idle time. Therefore, the optimal power allocation factor for this greedy scheme can be easily obtained by numerically solving the following one dimensional optimization problem: $\beta_{opt} = \arg \max_\beta \gamma_{su}(\beta)$, subject to $r_{pu}(\beta) \geq \xi r_{pu}^*/\xi C$. 

Figure 4.4: Throughput of the SU vs. the duty cycle $\xi$ for different $P_2$’s. Channel parameters: $P_1 = 1$, $a = 2$, $b = 1$, $c = 1$.

Figure 4.5: Power allocation factors vs. the duty cycle $\xi$ for the greedy power allocation. Channel parameters: $P_1 = 1$, $a = 2$, $b = 1$, and $c = 1$. 

\[\text{SU throughput (bps/Hz)}\]
\[\beta = 1, P_2 = 1\]
\[\beta_{opt}, P_2 = 1\]
\[\beta = 1, P_2 = 10\]
\[\beta_{opt}, P_2 = 10\]
4.3.3 Numerical Results and Discussions

Fig. 4.4 represents the ASE of the SU vs. the duty cycle $\xi$ for different SU powers and power allocations. As can be seen, the SU throughput decreases as the PU occupies the channel for more fraction of the time. However, even when the PU occupies the channel for 100% of the time, by utilizing the proposed scheme, the SU can achieve 14% of its point-to-point rate when $P_2 = 1$, or 37% of its point-to-point rate when $P_2 = 10$. It can also be observed that the gap between the optimal greedy power allocation and the selfless power allocation tend to vanish as the $P_2$ decreases.

Fig. 4.5 shows the optimal values of power allocation factors $\alpha$ and $\beta$ vs. the duty cycle $\xi$ when the greedy algorithm is utilized. $\alpha$ is chosen large enough to ensure successful decoding of the cooperative message at node 2, and $\beta$ obtained by numerically solving the optimization problem. As can be seen, when the duty cycle increases the optimal value for $\alpha$ decreases while that of $\beta$ increases. In other words, the SU tends to fully cooperate with the PU by assigning most of its power to forward the message of the PU. By adopting such a strategy, the PU can offload its data faster; therefore, the idle time, the time that the SU can fully occupy the channel, increases.

4.4 Chapter Summary

In this chapter, a simple cooperation scheme for the CoCR is introduced, and its achievable rate region is analytically derived. In this cooperation scheme, each user only has one power allocation parameter to tune. Moreover, two power allocation schemes (selfless and greedy) for the CoCR are proposed, and the optimal power allocation values for each scheme are obtained. The selfless algorithm is simpler in the sense that the power allocation factor $\beta$ is binary. Therefore, it incurs less overhead on the network. As observed, this power allocation method can result in a performance very close to that of the greedy one when the SU power is small.
Chapter 5

Clustering for Device-to-Device Assisted Virtual MIMO

In Chapters 3 and 4, we only considered the two-user case. In this chapter, we take a step towards scaling the network by considering a scenario with many users. Some of these users are sources (analogous to the PU) and some of the users are potential relays (analogous to the SU). The difference with previous chapters, however, is that the relays only decode and forward the message of the sources and do not have any message of their own. This assumption significantly simplifies the analysis. We also exploit beamforming and other MIMO techniques to improve the spectral efficiency of the sources.

More specifically, in this chapter, we treat the problem of clustering algorithms for multiple sources by utilizing idle UEs as assisting devices (relays) in a VMIMO setup with limited feedback. In the scenario under consideration, all UEs share the same channel, and they are spatially distributed according to a PPP. A source either transmits its message without forming a VMIMO system, or it can be clustered with other idle UEs in a VMIMO setup if the latter act improves its spectral efficiency. When a source participates in the VMIMO configuration, its message is conveyed to its serving AP in a two-phase DF relaying manner, and unlike some

\footnote{Contents of this chapter have been mainly published in [11] (Digital Object Identifier: 10.1109/TWC.2014.012114.120283).}
prior studies (e.g., [54, 55]), the direct link is also considered. Moreover, an approximate pre-
coding is performed by which the transmit signal of each UE in the second phase is multiplied
by a discrete unity complex number. Due to the limited feedback assumption, unlike [70], these
precoding weights are chosen from a finite cardinality codebook.

The rest of this chapter is organized as follows. In Section 5.2, we approximately solve
the NP-hard problem of precoding optimization by a combination of the search space reduc-
tion and SDR. In Section 5.3, we study a special case with a single source and analytically
derive an upper bound on the achievable spectral efficiency. This upper bound encompasses
the stochastic geometry of the problem as well as the randomness in channel gains due to the
log-normal shadowing. Then we propose a greedy algorithm that achieves this bound, and
hence, it is optimal in this case. We then leverage this knowledge in Section 5.4 to develop a
clustering algorithm for multiple sources. Our proposed algorithm is efficient in the following
senses: firstly, it is run in polynomial time; secondly, it eliminates the need for the backhaul
communication between APs since it is not required to control the leakage interference; thirdly,
it can achieve significant performance gains (cf. Fig. 5.3 Fig. 5.4). In Section 5.5, numerical
results are provided demonstrating, firstly, that the performance of our algorithm is very close
to that of an exhaustive clustering, and secondly, that our algorithm can improve the tradeoff
between the spectral efficiency and energy efficiency of the implementation.

5.1 System Model and Problem Statement

5.1.1 Distribution of UEs and Scheduling Model

We assume that UEs are randomly distributed in a two dimensional field according to a PPP
with rate \( \lambda \) per unit area. The Poisson assumption has been generally accepted as a proper
model for the spatial distribution of UEs in wireless networks specially in the presence of a
large population of users (cf. [102-104]).

Let \( \xi \) be a realization of the spatial distribution of UEs i.e., there are \( N_{UE}(\xi) \) UEs on the
Figure 5.1: Illustration of the system model for $M = 2$ destinations (APs) and 7 UEs ($N_{UE} = 7$). In this case, UEs 1 and 2 are scheduled to transmit (sources), and UEs 3 and 7 are assisting the UE 1 in the VMIMO configuration. Black lines represent transmission in the first phase, and green lines represent transmission in the second phase.

field given this realization. We further assume that $N_{UE}(\xi)$ remains stationary for a relatively long time. Therefore, we omit $\xi$ hearafter and let UEs be indexed as $1, 2, \cdots, N_{UE}$. Without loss of generality, we assume there are $M$ APs and the scheduler schedules one UE per AP, therefore, there are $M$ sources. Let these $M$ sources be indexed as $1, \cdots, M$ where $M \leq N_{UE}$.

### 5.1.2 Structure of the VMIMO and Corresponding Rates

It is assumed that each UE is equipped with one transmit antenna, and each AP is equipped with $N_{rx}$ receive antennas. Each source $s$ either directly transmits to its respective destination without the assistance of other relays or adopts a two-phase transmission where it is assisted by the set of idle relays $A_s$. In the latter case, in phase one, the source broadcasts its message (codeword), and all $k, k \in A_s$ decode the message of the source while the AP postpones the decoding. In phase two, the source $s$ and all the $k, k \in A_s$ transmit the same message...
as that of phase one with the precoding, and then, the AP decodes this message by augmenting received signals in phases one and two. In more precise words, if the source \( s \) is assisted by other UEs, it transmits the codeword \( x_j^{(1)} \) in phase one and repeats the same codeword in phase two. On the other hand, if the source \( s \) is not assisted by other UEs, it transmits \( x_j^{(q)} \) in phase \( q \) where \( q = 1, 2 \). We further assume that the symbols in the codeword are power normalized, i.e., \( E[|x_j|^2] = 1 \).

Let \( y_d^{(1)} \) be the received signal in phase one at a specific time at the AP \( d \). We have

\[
y_d^{(1)} = \sum_{j=1}^{M} h_{jd} \sqrt{P_j} x_j^{(1)} + n_d^{(1)}
\]

(5.1)

where \( y_d^{(1)} \) is an \( N_{rx} \times 1 \) vector, \( h_{jd} \) is an \( N_{rx} \times 1 \) vector whose elements are the channel gains between the UE \( j \) and the AP \( d \), \( x_j^{(1)} \) is the power normalized symbol transmitted by the source \( j \) in phase one, \( n_d^{(q)} \) is the AWGN vector in phase \( q \), and \( P_j \) is the transmit power of the UE \( j \). Also, let \( \mathcal{L} = \{1, \cdots, L\} \) be the subset of sources that are assisted in the described VMIMO setup. In the second phase, UEs that participate in VMIMO precode their transmit signal with a unity complex weight factor. Therefore, the received signal at the AP \( d \) in phase two can be written as

\[
y_d^{(2)} = \sum_{j=1}^{L} H_{jd} w_j x_j^{(1)} + \sum_{j=L+1}^{M} h_{jd} x_j^{(2)} + n_d^{(2)}
\]

(5.2)

where

\[ H_{jd} = \begin{bmatrix} h_{jd} \sqrt{P_j} & h_{k_1} \sqrt{P_{k_1}} & \cdots & h_{k_{|A_j|}} \sqrt{P_{k_{|A_j|}}} \end{bmatrix}, k_i \in A_j. \]

(5.3)

The column vector \( w_j \) contains \( |A_j| + 1 \) precoding weights \( w_{jk} \) where

\[ w_{jk} \in \{1, w, \cdots, w^{N_w-1}\}. \]
and $w$ is the $N_w$-th principal root of unity. The augmented received signal at the AP after two phases is

$$Y_d = \sum_{j=1}^{L} \left[ h_{jd} \sqrt{P_j} \right] x_j^{(1)} + \sum_{j=L+1}^{M} \left[ h_{jd} \sqrt{P_j} \right] x_j^{(1)} + \sum_{j=L+1}^{M} \left[ 0 \right] x_j^{(2)} + \left[ \frac{n_d^{(1)}}{\sqrt{P_j}} \right] \right]. \quad (5.4)$$

Employing a linear MMSE decoder [105], we first obtain the spectral efficiency of sources that are not assisted by relays. For the source $s$, $s \in \{L + 1, \cdots, M\}$, the capacity in phase $q = 1, 2$ can be computed as

$$c_s^{(q)} = C \left( P_s h_{sd}^{\dagger} K_z^{(q)} h_{sd} \right) \quad (5.5)$$

where

$$K_z^{(1)} = \sum_{j=1, j \neq s}^{M} P_j h_{jd} h_{jd}^{\dagger} + \sigma_N^2 I_{N_{rx}};$$

$$K_z^{(2)} = \sum_{j=1}^{L} H_{jd} w_j w_j^{\dagger} H_{jd}^{\dagger} + \sum_{j=L+1, j \neq s}^{M} P_j h_{jd} h_{jd}^{\dagger} + \sigma_N^2 I_{N_{rx}}.$$

The overall rate of the source $s$ is the average of its rates over two phases, i.e.,

$$r_s = \frac{1}{2} \left( c_s^{(1)} + c_s^{(2)} \right), \quad s \in \{L + 1, \cdots, M\}. \quad (5.6)$$

The aggregate capacity between the source $s$ when assisted by the relays $A_s$ in two phases and the destination $d$, can be written as

$$c_s = C \left( \bar{h}_{sd}^{\dagger} K_z^{-1} \bar{h}_{sd} \right) \quad (5.7)$$
where

\[
\hat{h}_{sd}^+ = \left[ h_{sd}^+ \sqrt{P_s} \ w_s^H H_{sd}^+ \right],
\]  
(5.8)

and

\[
K_z = \sum_{j=1, j \neq s}^L \hat{h}_{jd}^+ \hat{h}_{jd}^+ + \sum_{j=L+1}^M \begin{bmatrix} P_j h_{jd} h_{jd}^+ & 0 \\ 0 & P_j h_{jd} h_{jd}^+ \end{bmatrix} + \sigma_N^2 I_{2N_{rx}}.
\]  
(5.9)

Therefore, the overall spectral efficiency of the relay assisted source \( s \) can be written as

\[
r_s = \frac{1}{2} \min \left\{ \{ c_s \} \cup \{ r_{sl} : l \in A_s \} \right\}, \quad s \in \mathcal{L}
\]  
(5.10)

where \( r_{sl} \) is the rate between the source \( s \) and UE \( l \) (computed similar to \( c_s^{(1)} \)), and \( c_s \) is the rate in (5.7). The half factor is present since the assisted source \( s \) transmits a repeated version of its message in two consecutive time slots.

Fig. 5.1 illustrates the above described system model. In this figure, \( \mathcal{L} = \{1\} \), and \( A_1 = \{3, 7\} \).

### 5.1.3 Statement of the Problem

We seek to maximize the harmonic mean utility subject to disjoint set of assisting relays for each source and quantized unity precoding weights. This goal can be mathematically expressed as

\[
\max \frac{M}{\sum_{i=1}^M r_i^{-1}}
\]  
(5.10a)

s.t. \( w_{jk} \in \{1, w, \cdots, w^{N_w-1}\}, \ j \in \mathcal{L} \); \hspace{1cm} (5.10b)

\[ A_i \cap A_j = \emptyset, \quad i \neq j; \hspace{1cm} (5.10c) \]

\[ \bigcup_{j=1}^M A_j \subseteq \{M + 1, \cdots, N_{\text{UE}}\}. \hspace{1cm} (5.10d) \]
Remark 1 In a multi-source network, there are generally four utility functions that can be maximized: weighted sum-rate utility, proportional fairness utility, harmonic mean utility, and min-rate utility \[106\]. The last three utilities attain fairness in the network with different trade-offs between the aggregate throughput and fairness. In this thesis, we seek to maximize the overall spectral efficiency of the network given that each source is allowed to transmit the same number of information bits, and this objective leads to the harmonic mean utility. In more precise words, assume that each source \( i \) is allowed to transmit \( \kappa \) bits of information conveyed in the time interval \( T_i \) where \( \kappa = T_i R_i \). Therefore, the overall ASE is given as the ratio of the total number of bits to the total time, i.e., \( \kappa M / \sum_i T_i \). Simplification of this ratio leads to the harmonic mean of individual rates in (5.10a).

5.2 An Approximation Method for the Precoding Problem

Precoding is performed when users are clustered together as a VMIMO device to improve the performance by forming a beam towards the desired receiver. This method is also known as the transmit beamforming. To achieve an optimal performance, the precoding problem and the clustering problem should be considered jointly. However, due to the exponential complexity of the joint problem, they are considered separately in this work.

To approach the NP-hard problem of precoding, we isolate the precoding search space for each VMIMO cluster (associated with one source) and show in the following that SDR can be applied to obtain an approximate solution.

The SINR term in (5.7) can be written as

\[
\tilde{h}_{sd}^\dagger K^{-1}_{a_s} \tilde{h}_{sd} = \left[ h_{sd}^\dagger \sqrt{P_s} \right] \left[ w_s^\dagger H_{sd}^\dagger \right] K^{-1}_{a_s} \left[ h_{sd} \sqrt{P_s} \right] \left[ H_{sd} w_s \right] = \tilde{w}_s^\dagger Q_s \tilde{w}_s \tag{5.11}
\]
where \( \tilde{w}_s = \begin{bmatrix} 1 & w_s^\dagger \end{bmatrix} \), and

\[
Q_s = \begin{bmatrix} h_{sd}^\dagger \sqrt{P_s} & 0 \\ 0 & H_{sd}^\dagger \end{bmatrix} K_{e_s}^{-1} \begin{bmatrix} h_{sd} \sqrt{P_s} & 0 \\ 0 & H_{sd} \end{bmatrix}.
\]

The maximization of the SINR in (5.11) is tantamount to the following maximization problem

\[
\text{max } \tilde{w}_s^\dagger Q_s \tilde{w}_s \tag{5.12a}
\]

s.t. \( \tilde{w}_s \in \{1, w, \ldots, w^{N_w-1}\} \). \tag{5.12b}

The optimization problem expressed in (5.12) is called the discrete complex quadratic optimization problem which belongs to the class of NP-hard problems [107]. However, it can be approximately solved by using the SDR method. Considering the equality

\[
\tilde{w}_k^\dagger Q_k \tilde{w}_k = \text{tr}(Q_k \tilde{w}_k \tilde{w}_k^\dagger),
\]

the SDR version of (5.12) can be written as

\[
\text{max } \text{tr}(Q_s W) \tag{5.13a}
\]

s.t. \( W \succeq 0 \), \tag{5.13b}

\[
W_{ii} = 1 \tag{5.13c}
\]

where \( W \) is a new variable for the relaxed optimization problem. There are efficient numerical solvers for this semidefinite (SD) problem that can solve it in polynomial time. In this thesis, we first use the CVX package [108] to solve this optimization problem.

Next step in solving the precoding problem is to approximate appropriate precoding weighs from the solution of (5.13). There are two categories of approximation methods to find the solution of (5.12) from that of (5.13). In the first category, randomized methods are utilized [107].
with an approximation bound guaranteed in average while in the second category, deterministic methods are applied \[109\]. We adopt a deterministic approximation method in this thesis since achieving the approximation bound in randomized methods requires a large number of random trials.

Let $W^{opt}$ be the solution to (5.13). This solution can be decomposed as

$$W^{opt} = \sum_{k \geq 1} v_k v_k^\dagger \nu_k$$  \hspace{1cm} (5.14)$$

where $\nu_k$’s are rank ordered eigenvalues of $W^{opt}$, and $v_k$ is the eigenvector corresponding to $\nu_k$. To approximate $\tilde{w}_s$ in (5.12), the eigenvector corresponding to the largest eigenvalue of $W^{opt}$, $v_1$, is chosen and quantized. Let $v_{1i}$ be the $i$-th element of the eigenvector $v_1$. Also, let $w'_s$ be a new vector with elements $w'_{si}$. For some integer $m = 0, \ldots, N_w - 1$, the quantized value $w^m$ is assigned to $w'_{si}$ if

$$\arg(v_{1i}) \in \left[(m - 1/2) \arg(w), (m + 1/2) \arg(w) \right].$$  \hspace{1cm} (5.15)$$

One more step is needed before these quantized values can be express as the solutions to (5.12). Since the first element in $\tilde{w}_s$ is unity, the $w'_{si}$’s need to be phase rotated accordingly such that the first element becomes unity. This rotation is simply performed by dividing each element of $w'_s$ by its first element $w'_{s1}$. Therefore, the final solution to the optimization problem in (5.12) can be written as

$$\tilde{w}_{si} = w'_{si} w'_{s1}. \hspace{1cm} (5.16)$$

Note that the overall sol of (5.12) is not affected by this phase rotation.

Assuming a limited feedback, each UE participating in the VMIMO setup receives the index of the precoding weight in the precoding codebook, \{1, $w, \ldots, w^{N_w-1}\}$, through some feedback mechanism. Since the size of the precoding codebook is $N_w$, this feedback requires
\[ \log_2(N_w) \text{ bits.} \]

### 5.3 Clustering for a Single Source

In this section we develop an optimal clustering algorithm when there is only one source \((M = 1)\) and one destination. This algorithm will be later used in Section 5.4 for the general case with multiple sources.

To make computations tractable, we temporarily assume that \(N_{rx} = 1\); therefore, the channel gain between users \(l\) and \(m\) can be expressed as \(h_{lm} = |h_{lm}| e^{j\theta_{lm}}\). We further assume that channel gains are influenced by the path loss (PL) and log-normal shadowing\(^2\), therefore,

\[ |h_{lm}|^2 = Gd_{lm}^{-\alpha} 10^{0.1\sigma_{dB}V_{lm}} \]  

(5.17)

where \(G\) is a constant depending on the operating frequency and the antenna gains, \(d_{lm}\) is the distance between users \(l\) and \(m\), \(\alpha\) is the path loss exponent, \(\sigma_{dB}\) is the shadowing dB-spread, and \(V_{lm}\) is a standard Gaussian random variable. In addition, it is assumed that UEs are power controlled and the received SNR at the AP \(d\) is constant, i.e.,

\[ |h_{ld}|^2 \frac{P_l}{\sigma_N^2} = \gamma, \quad l = 1, \cdots, N_{UE} \]  

(5.18)

### 5.3.1 An Upper Bound on the Average Spectral Efficiency

In order to evaluate any algorithm, a performance bound on the achievable spectral efficiency is needed. This performance bound is given in the following theorem.

\(^2\)Note that the channel gain is usually composed of three major components: path loss, slow fading (shadowing) modelled as a log-normal random variable, and fast fading modelled as a Rayleigh random variable. However, the mean channel gain is only affected by the path loss and shadowing.
Figure 5.2: An illustration of the spatial distribution of source (s), relays, and the destination (d). \(d_{\text{max}}\) is the maximum radios around the source where the relays are spatially distributed. In this example, the set of all available relays is \(\{1, \cdots, 7\}\), and the set of assisting relays is \(A_s = \{1, 2, 3\}\).

**Theorem 8** For the system described in this section, there exists an upper bound on the ASE (bps/Hz) of the source s. For \(\sigma_{dB} = 0\), this upper bound is given as

\[
C(\gamma) + \int_{C(\gamma)}^{\infty} \sum_{k \geq k_r} \frac{[\lambda \Psi_s(r)]^k}{k!} e^{-\lambda \Psi_s(r)} dr \tag{5.19}
\]

where \(\lambda\) is the average UE density (number of UEs per unit area), and \(\Psi_s(r)\) is the \(r\)-achievability area and given by

\[
\Psi_s(r) = \pi \left[ \frac{GP_s/\sigma_X^2}{2^{2r} - 1} \right]^{2/\alpha}, \tag{5.20}
\]

and \(k_r\) is the \(r\)-necessary number of relays and given by

\[
k_r = \left[ \sqrt{\frac{2^{2r} - 1 - \gamma}{\gamma}} - 1 \right]. \tag{5.21}
\]

**Proof:** See Appendix B.5.
There are a number of important implications associated with Theorem 8. Firstly, the second term \( \int C(\gamma)dr \) in \( (5.19) \) represents the average improvement in the spectral efficiency achieved by the formation of the VMIMO as compared with the baseline spectral efficiency \( C(\gamma) \) without VMIMO. In other words, if the source can not benefit from the assistance of other UEs, the second term in \( (5.19) \) will be zero. Moreover, when the mean channel gains are influenced by PL, the \( r \)-achievable area and \( r \)-necessary number of relays can be used to simplify the clustering in a practical implementation when the geographical location of UEs is known. We state these results in the following two corollaries. The proofs follow directly from the proof of Theorem 8 and we skip them.

**Corollary 1** Let \( R_s \) be the maximum ASE expressed in \( (5.19) \), and the channel gains are affected by the PL. The number of relays necessary to achieve \( R_s \) is given by the \( R_s \)-necessary number of relays, \( k_{R_s} \).

**Corollary 2** Let \( R_s \) be the maximum ASE expressed in \( (5.19) \), and the channel gains are affected by the PL. The expected area in which assisting UEs are located is bounded by a disk centred at UE \( s \) with radius \( \sqrt{\Psi_s(R_s)/\pi} \).

For the general case with the log-normal shadowing, i.e., \( \sigma_{dB} > 0 \), an asymptotic upper bound can be obtained. The derivation of this bound is rather involved; we leave the details of this derivation to Appendix B.6 and state the results in the following lemma.

**Lemma 1** For the system described in this section and some positive \( \delta \), assume that relays are located within a disk with radius \( d_{max} \) (\( d_{max} > \delta > 0 \)) centred at the source \( s \). Let the probabilities \( \pi_k(r, \delta) \) be calculated according to \( (B.20) \). When \( \delta \to 0 \), the ASE of the source \( s \) can be upper bounded as

\[
C(\gamma) + \int_{C(\gamma)}^\infty \left[ \sum_{k \geq k_r} \pi_k(r, \delta) \right] dr.
\]

\( (5.22) \)
Note that the mean of the shadowing term in (5.17) is greater than one, i.e., \( E[10^{0.1\sigma dB_{\text{Vlm}}} ] \geq 1 \). In other words, it is expected that the log-normal shadowing increases the connectivity of the network, consistent with the results shown in [110]. This results in the bound for the ASE in Lemma 1 being greater than that in Theorem 8. This difference is illustrated in Fig. 5.3 and as can be seen, the log-normal shadowing increases the ASE. It is also illustrated that these bounds are tight for infinite number of precoding weights.

### 5.3.2 An Optimal Relay Selection Algorithm

For a fixed realization of UEs, let \( s \) be the source, and let the set of available idle UEs (relays) be \( \mathcal{I} = \{1, \ldots, N_{\text{UE}}\} \setminus \{s\} \). Denote as Algorithm 1 an algorithm that can perform optimal clustering when \( N_w = \infty \). Under this condition, an optimal precoding can be performed by multiplying the signal \( x_l \) by the precoding weight \( e^{-j\theta_{ld}} \), thus cancelling the phase rotation caused by the channel.

There are four major steps in Algorithm 1. In the first step, the candidate relays are discovered. Candidate relays are the ones whose rate to the source is greater than twice as much as that of the source to the AP. In the second step, these candidate relays are sorted in a descending ranking based on their link rate to the source. In the third step, these ranked candidate relays are added to the cluster one by one if with each addition, the spectral efficiency of the source increases. In the last step, it is decided whether the VMIMO system is formed or the source transmits in a one-phase manner.

**Proposition 1** Algorithm 1 selects the best subset of UEs to cluster with the source that maximizes the spectral efficiency.

**Proof:** See Appendix B.7.

**Remark 2** Algorithm 1 can be considered to belong to the popular category of greedy algorithms [117], i.e., it chooses the best relay at each step. In fact, all algorithms (including the ones in [47, 48, 54, 55]) that sweep through a sorted list of items to select the best set of
Algorithm 1 Clustering for the system described in Sec. 5.3

1: \( r_s \leftarrow \mathcal{C} (\gamma); \)  \( \triangleright \) Link rate without VMIMO
2: \( \mathcal{E}_s \leftarrow \{ l : l \in \mathcal{I}, \ r_{sl} > 2r_s \} \)  \( \triangleright \) Candidate relay discovery
3: \( \mathcal{E}_s^{\text{sorted}} \leftarrow \) Sorted \( \mathcal{E}_s \) descending w.r.t. \( r_{sj}, j \in \mathcal{E}_s \)
4: \( \mathcal{A}_s \leftarrow \emptyset; r_{\text{old}} \leftarrow 0 \)
5: \textbf{for} \( j \) in \( \mathcal{E}_s^{\text{sorted}} \) \textbf{do}
6: \hspace{1em} \( \mathcal{A}_s \leftarrow \mathcal{A}_s \cup \{ j \} \)
7: \hspace{1em} \( r_{\text{min}} \leftarrow \min \{ r_{sl} : l \in \mathcal{A}_s \} \)
8: \hspace{1em} \( p(w) \leftarrow \| h_{sd} \|_2^2 \frac{P_s}{\sigma_N} + \frac{1}{\sigma_N} \left\| \sum_{i \in \{s\} \cup \mathcal{A}_s} h_{id} w_i \sqrt{P_i} \right\|_2^2 \)
9: \hspace{1em} \( \gamma_{\text{AP}} \leftarrow \max_w p(w) \)  \( \triangleright \) Precoding
10: \hspace{1em} \( r_{\text{new}} \leftarrow \frac{1}{2} \min \{ C(\gamma_{\text{AP}}), r_{\text{min}} \} \)
11: \hspace{1em} \textbf{if} \( r_{\text{new}} > r_{\text{old}} \) \textbf{then}
12: \hspace{2em} \( r_{\text{old}} \leftarrow r_{\text{new}} \)
13: \hspace{1em} \textbf{else}
14: \hspace{2em} \( \mathcal{A}_s \leftarrow \mathcal{A}_s \backslash \{ j \} \)
15: \hspace{1em} \textbf{end if}
16: \textbf{end for}
17: \textbf{if} \( r_{\text{old}} < r_s \) \textbf{then}
18: \hspace{1em} \( \mathcal{A}_s \leftarrow \emptyset \)  \( \triangleright \) No VMIMO
19: \textbf{end if}

items fall under the category of greedy algorithms. However, there are four major differences between these works and our work in terms of the selection algorithm: firstly, in our work an adaptive number of relays is selected whereas the number of relays is fixed in [55]; secondly, by considering the effect of the direct transmission, our algorithm chooses no relay when the source can not benefit from relays in terms of achievable rates whereas in [48] and [54], the assumption is that source has to select at least one relay; thirdly, unlike [54,55], Algorithm 1 is not required to sort all the relays by establishing a necessary condition for the candidate relays in the discovery step of the algorithm (Line 2), thus reducing the complexity; lastly, unlike [47] and [48], our greedy algorithm is optimal.

To the best of the author’s knowledge, the fastest relay selection algorithm in the scenario under consideration has the complexity of \( O(\sqrt{\text{UE}} \log \sqrt{\text{UE}}) \) [55]. In the following, we show that the discover step can reduce the average run-time to a linear run-time in the number of available UEs. Let \( T(\sqrt{\text{UE}}) \) be the run-time of the Algorithm 1 when precoding is excluded.
Table 5.1: Description of the legends in Fig. 5.3

<table>
<thead>
<tr>
<th>Label</th>
<th>Method</th>
<th>$\sigma_{dB}$</th>
<th>$N_w$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(i)</td>
<td>Lemma 1</td>
<td>8</td>
<td>N/A</td>
</tr>
<tr>
<td>(ii)</td>
<td>Algorithm 1</td>
<td>8</td>
<td>$\infty$</td>
</tr>
<tr>
<td>(iii)</td>
<td>Theorem 8</td>
<td>0</td>
<td>N/A</td>
</tr>
<tr>
<td>(iv)</td>
<td>Algorithm 1</td>
<td>0</td>
<td>$\infty$</td>
</tr>
<tr>
<td>(v)</td>
<td>Algorithm 1</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>(vi)</td>
<td>Algorithm 1</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

**Theorem 9** $E[T(N_{UE})] = \Theta(N_{UE})$.

**Proof:** It can be shown that the average number of eligible relays ($E[|E_s|]$) remains constant as the number of relays increases. Therefore, the rest of the algorithm is performed on a constant number of relays, $|E_s|$. However, a linear time is required to perform the discovery step which makes the overall average run-time linear, $\Theta(N_{UE})$. For a detailed proof, see Appendix B.8.

Fig. 5.3 compares the ASE (bps/Hz) vs. $\lambda$ (UEs/m$^2$) for different received SNRs and different schemes as described in Table 5.1. For this figure, Algorithm 1 is performed over 100 random trials of spatial distribution of UEs. As can be seen, the spectral efficiency gained by Algorithm 1 overlaps with the upper bound on the ASE given in Theorem 8 and Lemma 1 when exact phase matching is performed, i.e., $N_w = \infty$ for graphs labeled (ii) and (iv). This observation further corroborates the optimality of Algorithm 1 stated in Proposition 1.

The precoding for this simulation (in graphs (v)–(vi)) is performed as described in Section 5.2. It can be speculated that this approximate precoding method improves the performance as compared to the case without the precoding ($N_w = 1$). Furthermore, the improvement in the spectral efficiency due to the VMIMO is more significant for poor performing UEs. For instance, for $\lambda = 0.01$ UEs/m$^2$, the VMIMO algorithm can boost the spectral efficiency of UEs with $\gamma = -10$ dB more than 400% as compared with the baseline spectral efficiency of $C'(\gamma) = 0.14$ bps/Hz whereas this improvement is about 10% for UEs with $\gamma = 10$ dB.
Figure 5.3: Average spectral efficiency vs. density of the UEs for various received SNRs. Curves are grouped for each received SNR $\gamma$, and the labels in the legend are explained in Table 5.1. Note that $\lambda = 0$ is treated as a special case in which there is only one source and no other UEs, $d_{\text{max}} = 25m$ for (i) and (ii), and $\delta = 0.05$ for (i).

## 5.4 Clustering for Multiple Sources

In this section, we extend the algorithm developed in Section 5.3 to the general case with multiple sources and develop Algorithm 2. Algorithm 1 is used in Algorithm 2 as a sub-algorithm with the following modifications: firstly, $C(\cdot)$ is obtained by the general formula given in (5.6) and (5.7); secondly, the precoding is performed according to Section 5.2.

Unlike Algorithm 1, we were unable to prove an optimality bound for Algorithm 2. However, through extensive simulations, we later illustrate that Algorithm 2 exhibits a near optimal performance when compared to that of an exhaustive clustering algorithm as far as the har-
monic mean utility metric is concerned.

### Algorithm 2 Clustering algorithm for multiple sources

\[
\begin{align*}
\mathcal{M} &\leftarrow \{1, \ldots, M\} \\
\mathcal{I} &\leftarrow \{M + 1, \ldots, N_{\text{UE}}\} \quad \triangleright \text{Available idle UEs} \\
\mathcal{M}^{\text{sorted}} &\leftarrow \mathcal{M} \text{ sorted ascending with respect to } c_s, s \in \mathcal{M} \\
\text{for } s \text{ in } \mathcal{M}^{\text{sorted}} \text{ do} & \quad \text{Perform Algorithm 1 for UE } s \text{ and idle UE set } \mathcal{I} \\
\mathcal{I} &\leftarrow \mathcal{I} \setminus A_s \\
\text{end for}
\end{align*}
\]

To motivate why a simple source sorting results in a near optimal performance, it should be noted that one important property of the harmonic mean is that it is limited by the smallest term\(^3\). This property implies that an increase in the smallest term can potentially lead to a significant increase in the harmonic mean. Therefore, in Algorithm 2, underprivileged sources are served first so as to enable them to benefit from the most available resources.

From an algorithmic perspective, we can distinguish Algorithm 2 from previous works on DF relay selection for multiple sources regarding three aspects. Firstly, our algorithm selects an adaptive number of relays for each source while in [57] and [59], one relay is selected for each source. As we argued, enforcing the source to select relays is not always optimal in terms of the spectral efficiency. Secondly, even though Algorithm 2 is oblivious to the leakage interference, the users are assumed to share the same channel. To isolate the relay selection for each user, however, in [57] and [58], it is assumed that sources utilize orthogonal channels. Thirdly, in [58–60], conflicts in selection are resolved by deploying some variation of the message passing procedure [94] (also known as auction rounds). For instance, [58] and [59] require up to \( M \) and \( M - 1 \) rounds of iteration, respectively. This number can be even larger in [60] depending on system parameters. In other words, the number of iteration rounds scale with the number of sources, and therefore, these algorithms may not be implementable in a large scale.

---

\(^3\) To clarify, consider the following inequality that holds for positive numbers \( r_1, r_2, \ldots, r_M \):

\[
\frac{M}{\sum_{i=1}^{M} r_i^{-1}} \leq M \min_{1 \leq i \leq M} r_i.
\]
scenario where the channels are dynamically and rapidly changing. This problem is addressed in our work by prioritizing sources and requiring only one round of iteration. Consequently, we show that a simple and agnostic algorithm can perform nearly optimal when the harmonic mean utility metric is considered, and it may not be necessary to resort to a complex and time taking solution.

5.5 Numerical Results

5.5.1 Simulation Setup and Parameters

In our simulations, we use the 3GPP channel model for the indoor environment [112, Sec. A.2]. According to this model the PL (in dB) for 2 GHz carrier frequency is given by $103.4 + 24.2 \log(d)$ where $d$ is the distance in km. The slow fading is modelled by the log-normal shadowing with dB-spread $\sigma_{dB} = 8$. Antennas are assumed to be omnidirectional with 0 dB antenna gain. The fast fading is assumed to be Rayleigh fading modelled as a Gaussian complex random variable with variance $1/2$ per real dimension. The noise power is $\sigma^2_N = -101$ dBm for 20 MHz bandwidth. The maximum transmit power of $\text{UEs}$ is limited to $P_{\text{max}} = 20$ dBm. $\text{UEs}$ are assumed to be power controlled, and the power control algorithm is assumed to average out the effect of Rayleigh fading, i.e., its decisions are only based on the PL and shadowing. As a result of this power control, each $\text{UE}$’s power is adjusted such that the received power at the serving $\text{AP}$ is $-80$ dBm excluding the interference. The number of receive antennas at the $\text{AP}$ is $N_{rx} = 4$. The field size is 100m×100m. We assume that there are five $\text{APs}$ and therefore, at each instance, there are five sources. Furthermore, to simulate the environments with interference, we assume there is an aggressor network with average density of $10^{-3}$ sources per square meter with a similar power control algorithm. The clustering algorithm has no control over this aggressor network.

For the sake of a better visualization in our simulation results, we use the effective $\text{SINR}$ (in dB) instead of the spectral efficiency. In other words, if $r_k$ is the spectral efficiency of the
source $k$, the corresponding effective $\text{SINR}_{\text{eff}}$ is calculated as

$$\text{SINR}_{\text{eff}} = 10 \log_{10} \left( 2^{r_k} - 1 \right).$$

For each $\lambda$ or $N_w$, the reported simulation results are averaged over 1000 trials. These trials encompass the random spatial distribution of UEs and APs, random log-normal shadowing, random Rayleigh fading, and random scheduling.
Figure 5.5: Average number of assisting UEs (relays) using Algorithm 2 vs. the baseline $\text{SINR}_{\text{eff}}$ of sources without VMIMO. $\lambda$ represents the UE density (UEs/m$^2$), and $N_w$ represents the number of precoding choices (there is no precoding for $N_w = 1$).

### 5.5.2 Discussion

The performance of Algorithm 2 is illustrated in Figs. 5.4, 5.5, 5.6, Table 5.2, and Table 5.3. In these illustrations, $\lambda$ represents the density of the network (UEs/m$^2$), and $N_w$ represents the number of precoding choices (there is no precoding when $N_w = 1$). For low densities of UEs, these performances are compared to that of an exhaustive clustering. However, since the run time of the exhaustive clustering grows exponentially with the density of UEs, for high density networks, it is not computationally feasible to run the exhaustive algorithm.

Fig. 5.4 shows the improvement in the $\text{SINR}_{\text{ eff}}$ of the sources after the formation of the VMIMO system vs. the baseline $\text{SINR}_{\text{eff}}$ of sources without the VMIMO. As can be seen, the
Figure 5.6: Cumulative distribution function of the $\text{SINR}_{\text{eff}}$ for different densities of UEs and precoding resolutions. $\lambda$ represents the UE density (UEs/m$^2$), and $N_w$ represents the number of precoding choices (there is no precoding for $N_w = 1$).

The spectral efficiency of poor performing sources is significantly increased whereas the spectral efficiency of the strong sources with high $\text{SINR}_{\text{eff}}$ is slightly degraded. For instance, when there is one UE in every 15 m$^2$ on average ($\lambda = 0.064$), sources with $\text{SINR}_{\text{eff}} = -10$ dB can gain more than 7 dB improvement in their $\text{SINR}_{\text{eff}}$ while sources with $\text{SINR}_{\text{eff}} = 10$ dB suffer less than 1 dB. This compromise, nevertheless, leads to an overall improvement (cf. Table 5.2).

Fig. 5.5 illustrates the average number of assisting UEs (relays) due to the VMIMO vs. the baseline $\text{SINR}_{\text{eff}}$ of sources without the VMIMO. As can be seen, the number of assisting UEs tend to increase when the precoding is performed because the aggregate uplink rate in (5.7) is increased due to this precoding. Hence, considering (5.10), the number of assisting relays can be increased.
Table 5.2: Improvement in the harmonic mean of the spectral efficiency for the coverage threshold $\text{SINR}_{\text{eff}} = -10 \, \text{dB}$.

<table>
<thead>
<tr>
<th>$\lambda$</th>
<th>Algorithm 2 $N_w = 1$</th>
<th>$N_w = 2$</th>
<th>$N_w = 64$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.002</td>
<td>2%</td>
<td>4%</td>
<td>4%</td>
</tr>
<tr>
<td>0.008</td>
<td>13%</td>
<td>18%</td>
<td>18%</td>
</tr>
<tr>
<td>0.064</td>
<td>43%</td>
<td>57%</td>
<td>58%</td>
</tr>
</tbody>
</table>

Table 5.3: Increase in the average consumed energy per bit $E_b \,(J/b)$ due to the VMIMO setup.

<table>
<thead>
<tr>
<th>$\lambda$</th>
<th>Algorithm 2 $N_w = 1$</th>
<th>$N_w = 2$</th>
<th>$N_w = 64$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.002</td>
<td>0.2%</td>
<td>-6%</td>
<td>-6%</td>
</tr>
<tr>
<td>0.008</td>
<td>11%</td>
<td>3%</td>
<td>3%</td>
</tr>
<tr>
<td>0.064</td>
<td>81%</td>
<td>69%</td>
<td>68%</td>
</tr>
</tbody>
</table>

Fig. 5.6 demonstrates the CDF of the $\text{SINR}_{\text{eff}}$ for different densities of UEs and precoding resolutions. As can be seen, the shift of the $\text{SINR}_{\text{eff}}$ towards higher values on the right, increases as the density of UEs increases. Also, the results clearly illustrate gains of up to 7 dB for the worst 5-10% of users and a gain of 3 dB on average.

The effect of the precoding codebook size ($N_w$) is illustrated in Tables 5.2 and 5.3. In our simulation, before implementing the VMIMO, the overall harmonic mean of the spectral efficiency is 0.83 (bps/Hz), and the $E_b$ is 69.5 $\mu$J/b. As can be seen, one bit of feedback is enough to provide the majority of the gains. Also, juxtaposing Table 5.2 and Table 5.3 reveals compelling results: firstly, by applying Algorithm 2, the spectral efficiency and energy efficiency are both improved for low densities of UEs; secondly, for higher densities of UEs, the precoding can significantly increase the spectral efficiency while reducing the average energy consumption per information bit.
5.6 Chapter Summary

In this chapter, we developed an efficient clustering algorithm for the D2D-assisted VMIMO systems with the limited feedback, and investigated the effect of the approximate precoding (transmit beamforming) and the UE density on the performance of this VMIMO system. As observed in the numerical simulations, the VMIMO system can significantly boost the performance of weak sources while only slightly degrading that of strong ones, thus leading to a considerable overall performance increment. These observations would suggest an approach of focusing the user clustering on weak users and not necessarily on strong users. In addition, it was shown that a single bit of feedback for the precoding weight is sufficient to provide the majority of the gain (Tables 5.2 and 5.3).
Chapter 6

Iterative Beamforming for D2D Assisted VMIMO

In Chapters 3 and 4 we studied two-user cases. Later in Chapter 5 we stepped towards scaling the network by considering multiple users with the simplifying assumption that relays do not have any message of their own. In this chapter, we take a step further by assuming that all the users have their own messages to convey. Similar to previous chapter, in this chapter, we also consider the problem of clustering and beamforming for D2D assisted virtual MIMO. However, there are two major differences with previous chapter in terms of the system model: firstly, all the users are sources (they have a message to transmit), and secondly, users can leverage a second radio access technology (RAT$_2$) to exchange information between each other. Note that the assumption of a RAT$_2$ is a reasonable assumption since nowadays most UEs operating on the licensed band (e.g., LTE bands) are equipped with transmitters and receivers for the Wi-Fi (2.4 GHz or 5 GHz) or Wi-Gig (60 GHz) unlicensed bands.

In terms of beamforming the approach of this chapter is different from that in the previous chapter in the following sense: In the previous chapter, the beamforming is performed for each VD individually using a feedback channel, and the beamforming weights only amended the transmission phase while keeping the power intact. In this chapter, however, the beamform-
ing is performed jointly for all the VDs using iterative updates on reciprocal channels, and the beamforming weights also adjust the transmit power. As it will be shown later, the joint beamforming tend to show a superior performance at the cost of the iteration overhead.

6.1 System Model and Problem Statement

In the scenario considered in this chapter, there are a number of co-channel UEs and a number of APs operating under the following assumptions: Each UE is equipped with a single transmit antenna whereas APs have multiple antennas, and UEs are transmitting to the APs (uplink). Single antenna UEs can be clustered together to form a VD. When clustered, UEs can leverage a RAT\textsubscript{2} to share their uplink messages. When UEs form a VD, they transmit their messages using a precoder. The role of the precoder is to adjust the transmit power as well as the transmission phase so that a network sum utility metric (including the proportional fairness utility metric) is maximized. The objective is to cluster the UEs together and find the precoding weights to maximize a sum-utility metric while minimizing the feedback overhead for the CSI.

6.1.1 System Model

In this section, the system model under consideration is explained. Table 6.1 summarizes the definition of some of the variables used in the system model.

Let there be a set of APs $\mathcal{K}$, and a set of UEs $\mathcal{M}$, where UEs intend to transmit their messages to the APs (uplink) using a first radio access technology (RAT\textsubscript{1}). UEs are also equipped with a RAT\textsubscript{2} which enables them to communicate with each other. RAT\textsubscript{1} and RAT\textsubscript{2} can be different in various ways including the frequency spectrum they occupy. Although considered general in this chapter, RAT\textsubscript{1} and RAT\textsubscript{2} can be implemented under existing wireless standards including LTE and IEEE 802.11.

Each UE is equipped with a single antenna while AP $k$ is equipped with $N_k$ antenna(s). Single antenna UEs can be clustered together to form a VD with multiple antennas. UEs in
Figure 6.1: An exemplary spatial distribution of UEs and APs with clustered UEs. In this example, \( M = 8, K = 2, L = 5, c = [1 \ 3 \ 2 \ 2 \ 4 \ 4 \ 5], \) and \( a = [1 \ 1 \ 2 \ 2 \ 2 \ 1 \ 1 \ 2]. \)

Each VD exchange their uplink messages using the RAT. After this exchange each UE shares the same information as other UEs in the same VD.

**Definition 6.1.1** Let the UEs in the VD employ an arbitrary protocol to exchange their messages after which they all share the same information bits. Assume they consume \( T \) units of time and \( W \) units of bandwidth for this exchange. If after this exchanging process, they all share the same \( N \) information bits, the empirical intra-cluster rate for the VD is defined as

\[
\tilde{r}_{[x]} \text{ def } \frac{N}{TW}.
\]  

A constructive protocol to achieve the empirical intra-cluster rate is given in Appendix B.9.

Each VD is assigned to one AP, and the assignments are represented by complementary
Table 6.1: Definition of some of the system variables.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
</table>
| $\mathcal{M}$ | Set of UEs; $\mathcal{M} \overset{\text{def}}{=} \{1, \ldots, M\}$.
| $\mathcal{K}$ | Set of APs; $\mathcal{K} \overset{\text{def}}{=} \{1, \ldots, K\}$.
| $\mathcal{L}$ | Set of VDs; $\mathcal{L} \overset{\text{def}}{=} \{1, \ldots, L\}$.
| $c_m$ | Clustering vector; $c_m = l$: UE $m$ is assigned to VD $l$.
| $a_m$ | Assignment vector; $a_m = k$: UE $m$ is assigned to AP $k$.
| $I_l$ | Set of UEs which belong to VD $l$; $I_l \overset{\text{def}}{=} \{m : c_m = l\}$.
| $I_{l,m}$ | Set $I_l$ such that $m \in I_l$.
| $l_k$ | Index of the VD assigned to the AP $k$.
| $k_l$ | Index of the AP assigned to the VD $l$.

indicators $l_k$ and $k_l$. The received signal at AP $k$ can be written as

$$y_k = \sum_{m \in \mathcal{M}} h_{km} v_m s_m + n_k \quad (6.2)$$

where $y_k \in \mathbb{C}^{N_k \times 1}$ is the received signal at the AP $k$, $h_{km} \in \mathbb{C}^{N_k \times 1}$ is the channel gain from the UE $m$ to the AP $k$, $n_k \in \mathbb{C}^{N_k \times 1}$ is the white Gaussian noise, $v_m \in \mathbb{C}$ is the precoding (beamforming) weight, and $s_m \in \mathbb{C}$ is the transmit symbol of the UE $m$. Note that all the UEs in a VD transmit the same message symbol to the respective AP; therefore, $s_m = s_l$ for all $m \in I_l$. Let the augmented channel matrix between the VD $l$ and the AP $k$ be defined as

$$\tilde{H}_{kl} = \begin{bmatrix} h_{ki_1} & \cdots & h_{ki_{N_l}} \end{bmatrix}, \quad i_j \in I_l, \quad (6.3)$$

and let $v_l = \begin{bmatrix} v_{i_1} & \cdots & v_{i_{N_l}} \end{bmatrix}^T$ where $i_j \in I_l$, and $N_l = |I_l|$. Therefore, the received signal at the AP $k_l$ assigned to the VD $l$ can be written as

$$y_{kl} = \tilde{H}_{kl} v_l s_l + \sum_{l' \in \mathcal{L} \setminus \{l\}} \tilde{H}_{kl'} v_{l'} s_{l'} + n_{kl}. \quad (6.4)$$
Let the achievable uplink rate for VD\(_l\) be denoted by \(r_{UL}^l\). This rate can be obtained as

\[
    r_{UL}^l = \log_2 \left( 1 + \mathbf{v}_l^\dagger \mathbf{H}_{k_l}^\dagger \mathbf{K}_l^{-1} \mathbf{H}_{k_l} \mathbf{v}_l \right) \tag{6.5}
\]

where \(\mathbf{K}_l\) is the covariance matrix of the coloured noise, i.e.,

\[
    \mathbf{K}_l = \sum_{l' \in \mathcal{L} \setminus \{l\}} \mathbf{H}_{k_l}^\dagger \mathbf{v}_{l'}^\dagger \mathbf{v}_{l'} \mathbf{H}_{k_l} + \sigma^2_N I. \tag{6.6}
\]

Since the UEs employ a DF protocol, each UE is required to decode the message of other UEs in the VD. Therefore, the effective rate of the VD\(_l\), \(r_{eff}^l\), is the minimum of the empirical intra-cluster rate \(\tilde{r}_{[x_l]}\) and the uplink rate \(r_{UL}^l\). That is,

\[
    r_{eff}^l = \min \{ \tilde{r}_{[x_l]}, r_{UL}^l \}. \tag{6.7}
\]

Fig. 6.1 depicts an exemplary distribution of UEs and APs with clustering. In this example, there are 8 UEs, 2 APs, and 5 virtual devices. Also, VDs 1, 3, and 4 are assigned to AP 1, and VDs 2 and 5 are assigned to AP 2.

### 6.1.2 Problem Statement

The objective is to find the clustering and assignment combination that maximizes a sum utility metric subjected to individual UE power constraints. In more precise words, we seek to solve
the following problem:

$$\max \sum_{l \in \mathcal{L}} u(\gamma^\text{eff}_l) \quad (6.8a)$$

s.t. 

$$\|v_m\|^2 \leq P_m, m \in \mathcal{M}; \quad (6.8b)$$

$$|\mathcal{L}| \leq |\mathcal{M}|; \quad (6.8c)$$

$$\mathcal{I}_l \cap \mathcal{I}_k = \emptyset, l \neq k, l \in \mathcal{L}, k \in \mathcal{L}; \quad (6.8d)$$

$$c_m = l \rightarrow a_m = k_l. \quad (6.8e)$$

In (6.8), $u(\cdot)$ is a utility metric such as the weighted rate utility, proportional fairness utility, harmonic mean utility, or min-rate utility [106]. The constraints in (6.8) ensure that firstly, the power constraints are met, secondly, the clusters are disjoint, and thirdly, all the UEs in a VD are assigned to the same AP.

### 6.2 Distributed MSE based Beamforming

Assume that the clusters are formed and each cluster is assigned to an AP. That is, the clustering and assignment vectors, $c$ and $a$, are known. (In Section 6.3, we discuss how to determine $a$ and $c$.) The AP $k_l$ deploys a linear decoder, $g_l \in \mathbb{C}^{N_{k_l} \times 1}$ to decode the message of the VD $l$; therefore, the decoded message $\hat{s}_l$ is given as

$$\hat{s}_l = g_l^\dagger y_{k_l}. \quad (6.9)$$

Let $u(\epsilon_l)$ represent the utility metric of the VD $l$ with MSE $\epsilon_l = E[\|\hat{s}_l - s_l\|^2]$. Note that this is a valid assumption since most network utility metrics are functions of rates $R_l$'s, and rates are functions of the MMSE. Using a Taylor expansion, the sum utility maximization can

---

1 The uplink rate of the VD $l$ in (6.5) can also be written as $r^\text{UL}_l = \log_2(\epsilon^\text{MMSE}_l)$ where $\epsilon^\text{MMSE}_l$ is the MSE when a linear MMSE decoder is employed and is given as $\epsilon^\text{MMSE}_l = 1 - v_l^\dagger \hat{H}_{k_l}^\dagger g_l$. 

be approximately expressed as a weighted sum MSE minimization [89]. That is,

$$\max \sum_{l \in L} u(\epsilon_l) = \min \left\{ -\sum_{l \in L} u'(\epsilon_l^{\text{old}}) \epsilon_l - o(\epsilon_l^2) \right\} \quad (6.10)$$

where \(o(\epsilon_l^2)\) represents the higher order terms.

Therefore, the optimization problem in (6.8a) and (6.8b) can be translated to the design of the precoding weights \(\{v_m\}\) and the decoding vectors \(\{g_l\}\) such that the weighted MSE of this decoding process is minimized when weights are defined as \(w_l = -u'(\epsilon_l^{\text{old}})\). In other words, the objective is to solve the following sub-problem:

$$\min \sum_{l \in L} w_l \epsilon_l, \quad (6.11a)$$

s.t. \(\|v_m\|^2 \leq P_m, \quad m \in M, \quad (6.11b)\)

\(c, a\) given. \( (6.11c)\)

**Proposition 2** The following precoding and decoding weights achieve a (local) optima for the problem in (6.11):

$$v_m = (\xi_m + \lambda_m)^{-1} \left( h_{am}^\dagger g_{am} w_{am} + \psi_m \right), \quad (6.12)$$

$$g_l = \left( \sum_{l' \in L} \tilde{H}_{l'lt} v_{l'} v_{l'}^\dagger \tilde{H}_{l'lt}^\dagger + \sigma_N^2 I \right)^{-1} \tilde{H}_{llt} v_l \quad (6.13)$$

where

$$\xi_m = \sum_{l \in L} h_{klm}^\dagger g_l w_l g_l^\dagger h_{klm}, \quad (6.14)$$
$\psi_m = \sum_{l \in \mathcal{L}} \sum_{i \in \mathcal{I}_m \setminus \{m\}} h_{k_i}^{\dagger} g_i w_l g_l^{\dagger} h_{k_i} v_i$, \hspace{1cm} (6.15)

$\lambda_m = \left[ \frac{-\kappa_1 + \sqrt{\kappa_1^2 - 4\kappa_2}}{2} \right]^+$, \hspace{1cm} (6.16)

and

$\kappa_1 = \xi_m + \xi_m^{\dagger}$, \hspace{1cm} (6.17)

$\kappa_2 = \|\xi_m\|^2 - \frac{\|h_{a_m}^{\dagger} g_{a_m} w_{a_m} + \psi_m\|^2}{P_m}$. \hspace{1cm} (6.18)

**Proof:** See Appendix B.10.

The updates in the Proposition 2 are obtained by applying the Karush-Kuhn-Tucker (KKT) conditions to the optimization problem in (6.11) as detailed in the Appendix B.10. A similar approach is used in \[82\] and \[86\]. Prior to the current work, since all transmit antennas belong to the same physical device, the optima of the Lagrangians are calculated when all transmit antennas are considered together. This means that to calculate the precoding weights, the channel to all the antennas on the device should be known which is a valid assumption given that all the antennas are on the same physical device. However, when the device is virtually formed, this condition can not be practically met. To address this challenge, an asymmetric approach is adopted in which we jointly optimize for the receiver but individually optimize for the transmit precoders. However, because of the $\psi_m$ term in (6.15), each UE should partially know the CSIs from the APs to other UEs belonging to the same VD. In the implementation of the algorithm, we ignore this term ($\psi_m \leftarrow 0$), and through simulations, we show that the performance loss is insignificant.
Algorithm 3 Iterative Beamforming

1: \( c \) and \( a \) are given
2: \( v_m^{\text{init}} = \sqrt{P_m} \) \( \triangleright \) precoding weights initialization
3: repeat
4: for all \( l \in L \) do
5: Update decoding vectors, \( g_l \), based on (6.13)
6: end for
7: for all \( m \in M \) do
8: Update \( \lambda_m \) and \( v_m \) based on (6.16) and (6.12)
9: end for
10: until max. number of iterations (\( \eta_{\text{IBF}} \)) reached

Algorithm 3 represents the decoding and precoding updates based on the Proposition 2. The iterative beamforming is performed until a certain number of iterations (\( \eta_{\text{IBF}} \)) is reached. In this algorithm, lines 4 to 6 correspond to the processing in the APs. In other words, each AP updates the decoder \( g_l \) corresponding to VD \( l \) assigned to it. Moreover, lines 7 to 9 correspond to the processing in the UEs.

6.3 Clustering

6.3.1 Proximity Measure

Defining a proper PM is a key factor in the performance of any clustering algorithm. The PM between VD \( i \) and VD \( j \), \( \mu_{ij} \), must be an indication of how suitable these VDs are to be clustered together, or in other words, it should indicate how close they are. In this section, various PMs are introduced and examined. Nonetheless, as argued in the introduction, the choice of the PM tends to be heuristic and is subjective to the problem at hand.

There is a necessary condition for a cluster to be viable. It is desired that after the clustering and beamforming, the uplink rate of the VDs \( i \) and \( j \) combined, \( r_{i,j}^{\text{UL}} \), is greater than the intra-cluster rate, i.e.,

\[
\frac{\bar{r}_{[I_i \cup I_j]}}{r_{i,j}^{\text{UL}}} \geq 1. \quad (6.19)
\]
Figure 6.2: An example run of Algorithm 4. These figure show the evolution of clusters as the algorithm advances. In these figure, black squares represents APs, and blue circles represent UEs. Green lines between UEs and APs show the UE-AP assignment. The red lines between the UEs indicate that corresponding UEs are clustered.

In other words, the communication between the UEs within a VD should not be the bottleneck. Therefore, the ratio in (6.19) can be used as a benchmark for the clustering in the following manner: since before merging VDs $i$ and $j$, the joint uplink rate is not available, this rate is approximated by the sum of individual uplink rates, i.e.,

$$ r_{i,j}^{UL} \approx r_{i}^{UL} + r_{j}^{UL}. $$

(6.20)
After approximation, the following ratio can be used as a rate-based PM

$$\mu_{ij}^{\text{rate}} = \frac{\hat{r}_{[I_i \cup I_j]}^{UL}}{r_i^{UL} + r_j^{UL}}. \quad (6.21)$$

The issue with this approach is that these proximity measures are based on the instantaneous rates and after each individual cluster update all the PMs throughout the network are affected, and hence, they need to be updated as well.

A second approach to define the PMs is to consider only the long term channel power gains (PL and shadowing). Intuitively, the PM should increase if the VDs are close to each other, and therefore, the PM should be proportional to the UE-to-UE channel gains. The minimum of the UE-to-UE channel gains can be used as a representative for the proximity of the UEs. On the other hand, the PM should decrease if the UEs are close to the serving AP. Considering these factors, we propose a max-min PM $\mu^{\text{max-min}}$, as follows

$$\mu_{lk}^{\text{max-min}} = \left( \max_{i \in I_l \cup I_k} \{ g_i \} \right)^{-1} \min_{i,j \in I_l \cup I_k} \{ g_{ij} \}$$

$$= \min_{i \in I_l \cup I_k} \{ g_i^{-1} \} \min_{i,j \in I_l \cup I_k} \{ g_{ij} \} \quad (6.22)$$

where $g_{ij}$ is the average channel power gain between UEs $i$ and $j$. In this definition of the PM, the term $\min_{i,j \in I_l \cup I_k} \{ g_{ij} \}$ pertains to the most far apart pair of UEs, and the term $\max_{i \in I_l \cup I_k} \{ g_i \}$ shows the maximum channel gain between the UE and the corresponding AP. Moreover, the size of the VD can also be taken into account, and a $\mu^{\text{max-min-size}}$ can be defined as

$$\mu_{lk}^{\text{max-min-size}} = |I_l \cup I_k|^{-1} \min_{i \in I_l \cup I_k} \{ g_i^{-1} \} \min_{i,j \in I_l \cup I_k} \{ g_{ij} \}. \quad (6.23)$$

Another approach to define the PM is to pair unbalanced UEs together. That is, the users with strong channel to the AP are paired with the ones with weak channel to the AP. To achieve this goal, the differential PMs $\mu^{\text{diff}}$ are defined to be proportional to the difference of the gains,
\[ \mu_{lk}^{\text{diff}} = \left| \log \max_{i \in I_l} \{ g_i \} - \log \max_{i \in I_k} \{ g_i \} \right| + \log \min_{i, j \in I_l \cup I_k} \{ g_{ij} \}. \] (6.24)

Similar to (6.22), the last term in (6.25) is to prevent the VD to become relatively large. In addition, similar to \( \mu^{\text{max-min-size}} \), the size of the VD can be considered and another differential PM can be defined as

\[ \mu_{lk}^{\text{diff-size}} = -\log |I_l \cup I_k| + \left| \log \max_{i \in I_l} \{ g_i \} - \log \max_{i \in I_k} \{ g_i \} \right| + \log \min_{i, j \in I_l \cup I_k} \{ g_{ij} \}. \] (6.25)

The performance of these five definitions of the PM are illustrated later in Section 6.4. As expected, the first PM, \( \mu^{\text{rate}} \), shows the best performance. Among the channel gain based PMs, the simple max-min PM exhibits a considerably close performance to that of the \( \mu^{\text{rate}} \).

### 6.3.2 A Greedy Clustering Algorithm

Greedy algorithms \[111\] are a popular class of algorithms that usually exhibits a significant (and sometimes optimal) performance with a low complexity. Greedy algorithms have been widely used for cooperative node selection \[47,48,54,55\]. In all these examples, the algorithm selects the best item at each step hoping that it leads to a global optimal choice. Similarly, Algorithm 4 represents a modified greedy clustering algorithm for the defined PMs. In this algorithm, \( U(\cdot) \) represents a utility function, and for the purpose of the simulations, a sum rate utility function is adopted here, i.e.,

\[ U(a, c, \{ v_m \}) = \sum_{i \in \mathcal{C}} r_i^{\text{eff}}. \] (6.26)

The algorithm attempts at most \( \eta_{\text{max}} \) trials of incremental updates of the clusters, and for each trial a greedy choice is made. That is, for each trial, the pair of VDs with the \( \kappa \)-th highest PM are chosen to be merged together. After this step, an IBF (Algorithm 3) is run, and if
the utility $U(\cdot)$ is not increased, the merge is discarded, and the algorithm looks for the VDs corresponding to the next best PM.

**Algorithm 4 IGU Clustering**

1: initializing: $\kappa \leftarrow 1$, $\eta \leftarrow 0$, $U^{\text{old}} \leftarrow 0$
2: while $\eta \leq \eta_{\text{max}}$ do
3: $\eta \leftarrow \eta + 1$
4: merge two VDs corresponding to the $\kappa$-th largest PM
5: IBF (Algorithm 3 / MWSMSE / max-SINR)
6: if $U(a, c, \{v_m\}) > U^{\text{old}}$ then
7: $U^{\text{old}} \leftarrow U(a, c, \{v_m\})$; $\kappa \leftarrow 1$
8: update PMs based on the new clustering
9: else
10: discard the merge
11: $\kappa \leftarrow \kappa + 1$
12: if $\kappa > \text{number of VD pairs}$ then
13: exit
14: end if
15: end if
16: end while

### 6.4 Numerical Results

In this section, numerical simulations are presented to compare the performance of our beamforming method with those in the literature. It is noteworthy that to the best of the author’s knowledge, there is no predecessor which directly considers the problem at hand. Therefore, we compare our results with the well known max-SINR algorithm [78], and the MWSMSE based algorithms, e.g., [82–84]. Also, it should be noted that the current state of the art iterative beamforming algorithms (cf., [89]) are not directly applicable to our scenario since they consider sum-power constraint whereas the problem in our scenario is constrained by per-antenna power constraint since UEs cannot share power.

In order to make the comparison possible, we modified the MSE based beamforming updates to satisfy the per-antenna power constraint. The details are presented in Appendix B.11, but the major challenge is that Lagrangians must be obtained through solving a system of
non-linear equations. Moreover, to attain some level of fairness, we employ the proportional fairness weighting (PFW) in the MSE base algorithms as well as our algorithm.

6.4.1 Simulation Setup

The channels are assumed to be AWGN where the noise is modelled with a circular symmetric Gaussian random variable with noise power $\sigma_N^2$ where $\sigma_N^2 = -101$ dBm for 20 MHz bandwidth. The channel power gain between users $i$ and $j$ are modelled as

$$g_{ij} = Gd^{-\alpha}10^{0.1\sigma_{dB}V}$$

where $G$ is a constant, $d$ is the distance, $\alpha$ is the path loss exponent, $\sigma_{dB}$ is the log-normal shadowing $dB$-spread, and $V$ is a normal random variable. The fast fading is assumed to be Rayleigh fading modelled as a Gaussian complex random variable with variance $1/2$ per real dimension. The rest of common simulation parameters are given in Table 6.2. Simulations are performed for $\eta_{user\ drop}$ random trails of physical user drop and log-normal shadowing. For each user drop, there are $\eta_{fading}$ trails of Rayleigh fadings.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>field dim.</td>
<td>$100 \times 100$ m$^2$</td>
<td>$P_m$</td>
<td>20 dBm</td>
</tr>
<tr>
<td>$B_{RAT1}$</td>
<td>20 MHz</td>
<td>$N_k$</td>
<td>4</td>
</tr>
<tr>
<td>$B_{RAT2}$</td>
<td>200 MHz</td>
<td>$\sigma_{dB}$</td>
<td>8</td>
</tr>
<tr>
<td>$f_{c_{RAT1}}$</td>
<td>2 GHz</td>
<td>$K$</td>
<td>6</td>
</tr>
<tr>
<td>$f_{c_{RAT2}}$</td>
<td>60 GHz</td>
<td>$\eta_{IBF}$</td>
<td>10</td>
</tr>
</tbody>
</table>

6.4.2 Discussions

Figs. 6.3, 6.4, and 6.5 illustrate various performance aspects of our algorithm. In these figures the cumulative distribution function (CDF) for (a) average cluster size, (b) the sum-rate, and
(c) the Jain fairness index \([113]\) is depicted. The Jain fairness index can be calculated assuming that all UEs in a VD transmit the same amount of information bits, i.e.,

\[
J = \frac{\left( \sum_{m \in M} r_m \right)^2}{M \sum_{m \in M} r_m^2} = \frac{\left( \sum_{l \in L} r_{\text{eff}}^l \right)^2}{M \sum_{l \in L} \left( \frac{r_{\text{eff}}^l}{N_l} \right)^2}.
\] (6.28)

Fig. 6.3 illustrates the performance of various iterative beamforming algorithm in the literature when compared to that of this work. As a baseline for the comparison, the max-SINR algorithm is performed on singleton clusters (no clustering). For the rest of graphs, Algorithm 4 is performed when various IBFs (including Algorithm 3, MWSMSE and max-SINR) are substituted in Line 5. Note that when MWSMSE or max-SINR is performed for clustered UEs, it is assumed that a genie informs each UE within a VD about the CSIs of other UEs in that VD; therefore, obviously, those genie-aided algorithms are expected to outperform Algorithm 3 in which beamforming updates are only based on local CSIs. Nonetheless, Algorithm 3 can improve both fairness and sum-rate when compared with no clustering (singleton clusters). Moreover, the performance of all the considered IBF algorithms increases when the rate-based PM is employed.

Fig. 6.4 illustrates the performance of the Algorithm 4 for various PMs. As can be seen, firstly, including the size of the VD in the PM tends not to change the performance. For instance, algorithms with \(\mu_{\text{max-min}}\) and \(\mu_{\text{max-min-size}}\) both show almost identical performances. Secondly, the max-min PMs (\(\mu_{\text{max-min}}\) and \(\mu_{\text{max-min-size}}\)) tend to exhibit a superior performance as compared to the differential PMs (\(\mu_{\text{diff}}\) and \(\mu_{\text{diff-size}}\)). These results seem to suggest to employ the \(\mu_{\text{max-min}}\) as the PM among the channel gain based PMs.

Fig. 6.5 compares the performance of the IBF for the case where only local CSIs are available (\(\psi_m \leftarrow 0\)) at each UE versus the case where each UE in the VD knows the CSI for other UEs (\(\psi_m\) is based on (6.15)). As can be seen, there is almost no degradation in the performance of our IBF due to the local knowledge of the CSIs. Also, as expected employing an instantaneous rate-based PM (\(\mu_{\text{rate}}\)) exhibits a superior performance compared with employing
the slagged max-min PM ($\mu^{\text{max-min}}$). Although juxtaposing Figs. 6.3 and 6.5 reveals that this performance superiority is insignificant when the number of clustering trials ($\eta_{\text{max}}$) are low.

6.5 Chapter Summary

In this chapter, the problem of beamforming and clustering for D2D assisted VMIMO is considered. In this scenario, single antenna UEs can be clustered together to form a VD. The UEs that belong to each VD first exchange their messages over the RAT. Having the same transmit signal, each UE in that VD transmits a weighted version of this signal. To find these weights, an iterative beamforming algorithm was proposed that composed of two stages in each round: After some initialization rounds, in the first stage, each UE calculates its beamforming weights based on the received signal from all the APs on the reciprocal channel and transmits some pilot signal to the APs. In the second stage, APs calculate their linear decoders and transmit back to the UEs using them as precoders. Unlike previous methods, the beamforming updates in this work only require local CSIs. Moreover, a greedy algorithm to cluster UEs was employed and various PMs were proposed. It was shown that the PM which only depends on long term channel gains performs very closely to that which depends on the expensive to compute instantaneous rates. Finally extensive simulations are performed, and the numerical results suggested that our method can improve both fairness and sum throughput even though it only uses local CSIs.
Figure 6.3: Comparison of performance of our algorithm (Algorithm 4) versus that in the literature. $\lambda = 3 \times 10^{-3}$, $\eta_{\text{user drop}} = 25$, $\eta_{\text{fading}} = 25$, $\eta_{\text{max}} = 50$. 
Figure 6.4: Illustration of the performance for different PMs. In Algorithm 3, local CSIs \((\psi_m \leftarrow 0)\) are assumed, and the weights \(w_l\) are chosen to maximize the proportional fairness. \(\lambda = 3 \times 10^{-3}\), \(\eta_{\text{user drop}} = 25\), \(\eta_{\text{fading}} = 25\), \(\eta_{\text{max}} = 25\).
Figure 6.5: Illustration of the effect of CSI knowledge on the performance of the proposed algorithm. In Algorithm 3, the weights $w_l$ are chosen to maximize the proportional fairness. $\lambda = 3 \times 10^{-3}$, $\eta_{\text{user drop}} = 25$, $\eta_{\text{fading}} = 25$, $\eta_{\text{max}} = 25$. 

(a) Average cluster size

(b) $\sum_{l} r_l^{\text{eff}}$ (bps/Hz)

(c) Jain fairness index for each UE

(d) Legends
Chapter 7

Conclusion

7.1 Thesis Summary and Concluding Remarks

In this thesis, we studied cooperative radio access architecture, protocols, and achievable bounds for the cognitive radio and device-to-device assisted virtual MIMO.

In Chapters 3 and 4, a two-user scenario was investigated in which the SU can transmit only if the spectral efficiency of the PU is not degraded. The goal of considering this scenario is to add a user to the system (the SU in this case) without any penalty to the current users (the PU) which is the promise of the CoCR concept. We started from a simple scenario in Chapter 3 where firstly, the SU transmits and receives in HD mode, and secondly, the messages of the PU and SU are not allowed to collide at the receivers. As shown, even with these simplifying and rather limiting assumptions, a CR can still operate, and there are power allocations that can achieve the goal. When the throughput (or the ASE) is concerned, the simple DF strategy without the SIC can perform as good as more complicated schemes with the SIC. In Chapter 4, we relaxed the second assumption so that the messages are allowed to collide at the receivers. Under this assumption, we showed that the DPC can be leveraged to enhance the spectral efficiency. Then we proposed a power allocation scheme in which each user had only one
In an effort towards scaling the wireless system, in Chapter 5 we considered a system with multiple users. In the scenario considered, there are multiple sources (analogous to the PU) and multiple potential relays (analogous to the SU). Each source tries to select the best subset of available UEs as relays so that its spectral efficiency is improved. The selected relays cooperate with the source in a HD and DF manner. It was observed that not any relay selection can improve the spectral efficiency since due to the repetition nature of the scheme, each source looses half of its spectral efficiency as soon as it adopts the two-phase cooperative transmission. This loss, however, can be compensated by SNR gains achieved by combing source and (proper) relay’s signals at the destination. It was shown that a threshold-based greedy algorithm can operate optimally with expected linear time (in the number of UEs) in the considered scenario. In the multi-source case, it was observed that an agnostic and simple algorithm can perform near optimal when the harmonic mean is considered as the network utility metric. The simulation results also showed that the spectral efficiency of the low SINR sources can significantly be increased whereas that of strong users may slightly be diminished which led to an overall utility increment. These results seem to suggest that the clustering is performed for low SINR sources rather than for high SINR ones.

We further generalized our assumptions in Chapter 6 where all the UEs are considered as sources and the goal is to maximize the overall network spectral efficiency (bps/Hz/UE) while achieving a level of fairness. To achieve this goal, instead of the selfish beamforming in Chapter 5, we adopted an iterative beamforming method in which all the UEs and APs are simultaneously involved. In this method, the beamforming vectors of each UE is chosen such that the UE maps its interference footprint at other non-intended APs to a subspace orthogonal to that of the intended signal at those APs. We obtained update rules that attempt to achieve this alignment of the interference by minimizing the weighted sum MSE. Unlike most previous

\footnote{We also relaxed the first assumption in Appendix C where the SU can simultaneously receive and transmit (FD).}
works, our updates only require local CSI at each UE. Then the performance of this beamforming method was illustrated in a clustering algorithm. The clustering algorithm incrementally merges a pair of most similar VDs at each step and performs the iterative beamforming. To indicated which UEs are suitable to be paired, different PMs are proposed. It was argued that a proper PM indicates the relative distance of UEs to the APs. Finally, extensive simulations were performed, and it was shown that our iterative beamforming updates with the max-min PM can outperform known results in terms of both the sum throughput and fairness.

7.2 Future Works

In the context of the proposed network access control to the problem of cooperative communications the following research directions are still open and further work can address these issues:

- In this thesis, we only considered the two-user case for the CR. The two-user case, however, can be extended to multi-user case in order to enable more users to utilize the spectrum. In the multi-user case, there are multiple PUs and multiple SUs each with their own messages. Note that the assumption here is that the performance of PUs are not degraded due to the addition of the SUs.

- In Chapter 4, we only considered the DF schemes in which the SU partially decodes the message of the PU. This assumption is valid when users are in the proximity of each other. However, investigating a AF schemes is also necessary to have a benchmark for comparisons.

- In the D2D assisted VMIMO only single antenna UEs are considered to be clustered together. The extension of this work to the case where UEs are equipped with multiple antennas remains as an interesting future work.

- For D2D assisted VMIMO in Chapter 6, only the repetition-based cooperative diversity
gain has been exploited; utilizing space-time-coded cooperative diversity or the multiplexing gain and studying the tradeoff between the diversity and multiplexing gains is a promising area for future investigations.

- In Chapters 5 and 6, we the UEs only cooperate with each other in a DF manner. Investigating a similar problem when UEs cooperate with each other in an AF manner remains as a demanding future work.
Appendix A

Perliminaries

In this chapter, some mathematical preliminaries are introduced. The introduced concepts include the capacity of a point to point channel, and MMSE decoders that are useful to understand the materials and proofs in this thesis.

A.1 Channel Coding Theorem

Throughout this thesis, the concept of the channel capacity is used numerous times. In this section, a brief introduction on the channel capacity theorem is provided.

A basic single user communication system model is illustrated in Fig. A.1. As can be seen, this model comprises a message $w$, an encoder which encodes the message set onto codeword $x^n$, a discrete memoryless channel (DMC) and a decoder that maps the channel outputs onto a message estimate. The channel is shown by $(\mathcal{X}, p(y|x), \mathcal{Y})$ where $\mathcal{X}$ is the set of channel input alphabets, $\mathcal{Y}$ is the set of channel output alphabets, and $p(y|x)$ is the channel transition probability function.

These concepts, however, can be found in the text books (e.g., [114] and [105]), and the familiar reader can skip them.
Figure A.1: Basic single user communication channel. The channel is represented by a conditional probability mass function $p(y^n|x^n)$. The encoder maps each message into a codeword $x^n$. Inversely, the decoder maps the channel output $y^n$ into a message estimate.

The channel input $X^n$ is subjected to a constraint

$$\frac{1}{n} \sum_{i=1}^{n} E \{ \phi(X_i) \} \leq \Gamma,$$  \hspace{1cm} (A.1)

where $\phi: \mathcal{X} \mapsto \{0\} \cup \mathbb{R}^+$ is the transmission cost function [116], $\mathbb{R}^+$ is the set of all positive real numbers, and $\Gamma > 0$ is a constant.

Being memoryless implies that $p(y^n|x^n) = \prod_{i=1}^{n} p(y_i|x_i)$. Furthermore, a $(2^{nR}, n)$ code for this channel can be defined as to include the following elements:

1. A set of messages $\mathcal{W} = \{1, 2, \cdots, 2^{nR}\}$. Throughout this work, we assume that the message $w$ is uniformly distributed on $\mathcal{W}$.

2. An encoding function that assigns a codeword $x^n(w)$ to each message $w$. The set of all codewords $\{x^n(1), x^n(2), \cdots, x^n(2^{nR})\}$ is called codebook $\mathcal{C}$.

3. A decoding function $g(\cdot)$ that maps the channel output $y^n$ onto the message set $\mathcal{W}$, i.e., $\hat{w} = g(y^n), \hat{w} \in \mathcal{W}$.

The rate $R$ of the code is defined as logarithm of the message size divided by the number of channel use. In other words,

$$R = \frac{\log_2 |\mathcal{W}|}{n} \text{ bits per channel use.}$$ \hspace{1cm} (A.2)

For above mentioned channel, let $\lambda_w = Pr\{\hat{w} \neq w | w \text{ sent}\}$ be the conditional probability of error given that the transmitted message is $w$. Then, the average probability of error $P_e^n$ for
A code \((2^nR, n)\) can be defined as

\[
P_e^{(n)} = \frac{1}{2^nR} \sum_{w=1}^{2^nR} \lambda_w. \tag{A.3}
\]

A rate \(R\) is said to be achievable if there is a code such that \(P_e^{(n)} \to 0\) as \(n \to \infty\). The supreme of all achievable rates is called the channel capacity.

**Theorem 10 (Shannon [115])** The capacity of a discrete memoryless channel is given by

\[
C = \max_{X \sim p(x)} I(X; Y). \tag{A.4}
\]

### A.1.1 Capacity of the AWGN Channel

When the noise is additive white Gaussian noise (which is the case in many practical communication channels), the alphabet sets of the channel input and channel output are assumed to be continuous.

For a point-to-point communication channel with additive white Gaussian noise, the channel output \(Y = X + Z\) where \(Z\) is an additive zero mean circularly symmetric Gaussian noise with variance \(N\), i.e., \(E[|Z|^2] = N\). The channel input \(X\) is a zero mean random variable with power \(E[|X|^2]\) limited to \(P\). The noise and the transmitted codeword are assumed to be independent. The capacity of this channel can be written as

\[
C = \log \left(1 + \frac{P}{N}\right). \tag{A.5}
\]

This expression is an important result that will be frequently used throughout this thesis.
A.2 Multi-Antenna Receivers with MMSE Decoder

Assume a scenario with multiple transmitters and a receiver with $N_{rx}$ receive antennas [105, Chapter 8]. The relationship between channel inputs and outputs can be written as

$$y = \sum_i h_i x_i + n$$  \hspace{1cm} (A.6)

where $y \in \mathbb{C}^{N_{rx} \times 1}$ is the received signal, $h_i \in \mathbb{C}^{N_{rx} \times 1}$ is the channel gain between transmitter $i$ and the destination, $x_i \in \mathbb{C}$ is the transmitted symbol of user $i$, and $n \in \mathbb{C}^{N_{rx} \times 1}$ is the circular symmetric Gaussian noise at the receiver. The receiver intends to decode the signal $x_k$. To find the optimal receiver, the interference plus noise can be considered as a non-white noise, and a linear decoder that minimizes the MSE can be applied. In other words, the received signal $y$ can be rewritten as the summation of the desired signal and a non-white noise $z = \sum_{i \neq k} h_i x_i + n$, i.e.,

$$y = h_k x_k + z.$$  \hspace{1cm} (A.7)

where $z$ is complex circular symmetric colored noise with an invertible covariance matrix $K_z$, $h$ is a deterministic vector, and $x_k$ is the unknown scalar symbol to be estimated. $z$ and $x_k$ are assumed to be uncorrelated. The covariance matrix $K_z$ can be obtained as

$$K_z = E[zz^\dagger]$$

$$= N_0 I + \sum_{i \neq k} P_i h_i h_i^\dagger$$  \hspace{1cm} (A.8)

where $N_0$ is the noise power, $I$ is an identity matrix, and $P_i$ is the transmit power of the $i$-th transmitter.

The received signal $y$ can be passed through the invertible linear transformation $K_z^{-1/2}$ such
that the noise $\tilde{z} = K_z^{-1/2} z$ becomes white:

$$K_z^{-1/2} y = K_z^{-1/2} h_k x_k + \tilde{z}. \quad (A.9)$$

Next, the output is projected in the direction of $K_z^{-1/2} h_k$ to get an effective scalar channel

$$(K_z^{-1/2} h_k)\hbar K_z^{-1/2} y = h_k^\dagger K_z^{-1} h_k x_k + h_k^\dagger K_z^{-1} z. \quad (A.10)$$

Hence, the corresponding achieved **SINR** is

$$P_k h_k^\dagger K_z^{-1} h_k \quad (A.11)$$

where $P_k$ is the transmit power of the signal $x_k$. Therefore, the rate at which the user $k$ can transmit is given as

$$R_k = \log_2 \left( 1 + P_k h_k^\dagger K_z^{-1} h_k \right). \quad (A.12)$$
Appendix B

Proofs

B.1 Proof of Theorem 1

Even though the CCR comprises multiple nodes, the point-to-point Shannon capacity results are applicable to it due to the fact that the collision is avoided. Consider (3.4) and (3.5): the PR forms a augmented codeword $Y^n_3$ which is a matrix with 2 rows and $n$ columns. The first row is the $n$-tuple received in the first time slot, and the second row is the $n$-tuple received in the second time slot. Note that symbol $i$ from the first time slot and the corresponding symbol $i+n$ from the second time slot are in the same column (column $i$), and $i = 1, \ldots, n$. Also, note that the message of the PT is encoded in the $n$-tuple $X^m_1$. According to the Shannon theory [115], in each channel use, the amount of information bits of $X^m_1$ that can be reliably decoded from $Y^n_3$ (when $n$ is large) is upper bounded by the mutual information between $X^m_1$ and $Y_3$, i.e., $I(Y_3; X^m_1)$. Here, there is a repetition: the same information bits are conveyed in the second time slot; therefore, the actual rate for the codeword $X^m_1$ is half of this mutual information.
term. In other words, under this scheme, the PR can decode $X'_1^n$ if $R_1 \leq 0.5 I(Y_3; X'_1)$.

$$R_1 \leq \frac{1}{2} I(Y_3; X'_1) = \frac{1}{2} (h(Y_3) - h(Y_3|X'_1))$$

$$= \frac{1}{2} h \left( \begin{bmatrix} \sqrt{P_1} X'_1 + Z_3^{(1)} \\ c \sqrt{\beta P_2} X'_1 + c \sqrt{\beta P_2} X'_1 + Z_3^{(2)} \end{bmatrix} \right)$$

$$- \frac{1}{2} h \left( \begin{bmatrix} Z_3^{(1)} \\ c \sqrt{\beta P_2} X'_1 + Z_3^{(2)} \end{bmatrix} \right)$$

$$= \frac{1}{4} \log \left| \begin{bmatrix} 1 + P_1 & c \sqrt{\beta P_1} P_2 \\ c \sqrt{\beta P_1} P_2 & 1 + c^2 P_2 \end{bmatrix} \right|$$

$$- \frac{1}{4} \log \left| \begin{bmatrix} 1 & 0 \\ 0 & 1 + c^2 \beta P_2 \end{bmatrix} \right|$$

$$= \frac{1}{4} \log \frac{1 + P_1 + c^2 P_2 + c^2 \beta P_1 P_2}{1 + c^2 \beta P_2}$$

(B.1)

where $I(\cdot; \cdot)$ represents the mutual information, and $h(\cdot)$ is the differential entropy (see [114] for definitions), and [114, Theorem 8.4.1] is used to compute (B.1).

Using a similar argument, let the SR form a augmented codeword $Y_4^n$. The SR can decode the message $X'_2^n$ if

$$R_2 \leq \frac{1}{2} I(Y_4; X'_2).$$

This mutual information can be computed, and (3.12) will be obtained. \qed

### B.2 Proof of Theorem 2

We imitate the proving technique in the Theorem 1. First, the decoders form augmented codewords $Y_3^n$ and $Y_4^n$. Then, the PR decodes the message of the SU reliably if $R_2 \leq 0.5 I(Y_3; X'_2)$. Having decoded $X'_2^n$, the PR can utilize it as a known information to decode the message of the
The message of the PU can be decoded without error if \( R_1 \leq 0.5I(Y_3; X_1' | X_2') \). Finally, the SR can decode the message \( X_2^m \) if \( R_2 \leq 0.5I(Y_4; X_2') \). Considering both constraints on \( R_2 \), clearly, \( R_2 \leq \min\{0.5I(Y_3; X_2'), 0.5I(Y_4; X_2')\} \) must be satisfied. □

### B.3 Proof of Theorem 3

A similar proof technique to that of the Theorem 1 is used. The PR can decode \( X_1^m \) if \( R_1 \leq 0.5I(Y_3; X_1') \). The SR can decode \( X_1^m \) if \( R_1 \leq 0.5I(Y_4; X_1') \). Having the message of the PT known, the SR can decode the message from the ST if \( R_2 \leq 0.5I(Y_4; X_2'|X_1') \). □

### B.4 Proof of Theorem 7

In this section, the proof of Theorem 7 is presented. In this proof, the notations \( X \overset{i.i.d.}{\sim} p(x) \) is used to indicate that the random variable \( X \) is drawn independent and identically distributed (i.i.d.) according to the probability measure \( p(\cdot) \) on \( \mathcal{X} \). All other notations follow the conventions in [114].

#### B.4.1 Codebook Generation

Let \( \mathcal{W}_e = \{1, 2, \ldots, 2^{nR_e}\} \) and \( \mathcal{W}_o = \{1, 2, \ldots, 2^{nR_o}\} \) be the set of messages for the PU. Also, let \( \mathcal{W}_2 = \{1, 2, \ldots, 2^{nR_2}\} \) be the set of messages for the SU. Furthermore, let \( w_e \in \mathcal{W}_e \), \( w_o \in \mathcal{W}_o \), and \( w_2 \in \mathcal{W}_2 \). The codebook is generated by the following rules:

- Generate \( |\mathcal{W}_e| \) codewords \( u^n \overset{i.i.d.}{\sim} \prod_{i=1}^n p(u_i) \) and index them as \( u^n(w_e) \);
- Generate \( 2^{nL} \) codewords \( v^n \overset{i.i.d.}{\sim} \prod_{i=1}^n p(v_i) \) and index them as \( v^n(l) \) where \( l = 1, 2, \ldots, 2^{nL} \);
- For each codeword \( u^n \), generate \( |\mathcal{W}_o| \) codewords \( x_o^n \overset{i.i.d.}{\sim} \prod_{i=1}^n p(x_{o,i}|u_i) \) and index them as \( x_o^n(w_o, w_e) \);
• For each codeword $u^n$, generate a codeword $x^n_e \sim \prod_{i=1}^{n} p(x_e | u_i)$ and index it as $x^n_e(w_e)$;

• For each codeword pair $(u^n, v^n)$, generate a codeword $x^n_2 \sim \prod_{i=1}^{n} p(x_2 | u_i, v_i)$ and index it as $x^n_2(l, w_e)$;

• The codewords $v^n$ are randomly and uniformly distributed into $2^{nR_2}$ bins;

• In the last step, the codebook is revealed to all transmitters and receivers.

This codebook generation is illustrated in Figure B.1

### B.4.2 Encoding and Decoding

A modified version of the block-Markov encoding [2] and a backward decoding [117] are used here. We consider 2 blocks of transmissions, and each block length is $n$.

In an odd block, node 1 transmits $x^n_o(w_o, w_e)$ where $w_o \in \mathcal{W}_o$ and $w_e \in \mathcal{W}_e$. In the next block, which is an even block, node 1 transmits $x^n_e(w_e)$. The encoding at node 2 is performed using the Gel’fand-Pinsker method. In an odd block, node 2 has an estimation of the message $\hat{w}_e$ transmitted by node 1. In the even block, when the message $w_2$ is about to be transmitted, node 2 seeks in the bin with index $w_2$ to find a codeword $v^n$ in that bin such that it is jointly typical with the codeword $u^n(\hat{w}_e)$. In other words, node 2 seeks for a message index $l$ such that $(v^n(l), u^n(\hat{w}_e)) \in T_{\epsilon}^{(n)}$, and $v^n(l)$ belongs to the bin $w_e$. The probability of not finding such a codeword $v^n$ vanishes as $n$ grows if $L - R_2 > I(U; V)$.

The decoding is performed differently for odd and even blocks. After receiving an odd block, node 2 declares the message $\hat{w}_e$ was transmitted if there is only one codeword $u^n(\hat{w}_e)$ such that $(u^n(\hat{w}_e); Y^n_{2,o}) \in T_{\epsilon}^{(n)}$. The probability of decoding at this step vanishes as $n$ grows if $R_e < I(Y_{2,o}; U)$.

Node 3 withholds performing the decoding until it receives the second block. After receiv-
ing $Y_{3,e}^n$ in the even block, node 3 first decode $\hat{w}_e$ if there is only one $\hat{w}_e$ such that

$$\left( u^n(\hat{w}_e), Y_{3,e}^n \right) \in T_e^{(n)}.$$

The probability of error at this step vanishes as $n$ grows if

$$R_e < I(Y_{3,e}; U).$$

Having decoded $w_e$, the decoder at node 3 declares $\hat{w}_o$ was transmitted if there is only one such $\hat{w}_o$ that satisfies

$$\left( u^n(w_e), x_o^n(\hat{w}_o, w_e), Y_{3,o}^n \right) \in T_e^{(n)}.$$

The probability of error at this step vanishes as $n$ grows if

$$R_o < I(Y_{3,o}; X_o | U).$$

The decoding at the node 4 is performed after the even block is received; the decoder declares that $\hat{l}$ was transmitted if there is only one codeword $v^n(\hat{l})$ such that

$$\left( v^n(\hat{l}), Y_{4,e}^n \right) \in T_e^{(n)}.$$

The probability of error at this step vanishes as $n$ grows if

$$L < I(V; Y_{4,e}).$$

As discussed, each message is being conveyed to the destination in two consecutive blocks. Therefore, the effective rate of the PU, $r_{pu} = \frac{1}{2} R_e + \frac{1}{2} R_o$, is

$$r_{pu} < \frac{1}{2} \left[ I(Y_{3,o}; X_o | U) + \min \{ I(Y_{2,o}; U), I(Y_{3,e}; U) \} \right]. \quad (B.2)$$
Similarly, since the encoders and decoders spend two blocks to decode $w_2$, the overall rate of the SU is half of the rate of $w_2$, i.e.,

$$r_{su} = \frac{1}{2} R_2 < \frac{1}{2} \left[ I(Y_{4,e}; V) - I(V; U) \right]. \quad (B.3)$$

### B.4.3 Gaussian Channel

To map the generated codebook into the Gaussian codebook, let the auxiliary random variables $U$, $\tilde{V}$, and $\tilde{X}_o$ as independent standard Gaussian random variables (zero mean and unity variance). It can be written

$$X_o = \sqrt{\alpha P_1} \tilde{X}_o + \sqrt{(1 - \alpha) P_1} U; \quad X_e = \sqrt{P_1} U; \quad (B.4)$$

$$X_2 = \sqrt{(1 - \beta) P_2} \tilde{V} + \sqrt{\beta P_2} U; \quad (B.5)$$

$$V = \sqrt{(1 - \beta) P_2} \tilde{V} + \lambda \left( b\sqrt{P_1} + \sqrt{\beta P_2} \right) U. \quad (B.6)$$

The parameters $\alpha$ and $\beta$ are power allocation factors and defined in the interval $[0, 1]$. The parameter $\lambda$ is employed to establish correlation between the random variables $U$ and $V$. The mutual information terms in Eqs. (B.2) and (B.3) can be obtained as follows: To compute the mutual information term $I(Y_{4,e}; V) - I(V; U)$, Costa’s DPC result [4] is used. The DPC is being performed at the SU in the even block knowing the $(b\sqrt{P_1} + \sqrt{\beta P_2}) U$ as the state of the channel. In this case the channel input is $\sqrt{(1 - \beta) P_2} \tilde{V}$, and the channel output is $Y_{4,e}$. Therefore, the choice of $\lambda = (1 - \beta) P_2/((1 - \beta) P_2 + 1)$ is optimal, and

$$I(Y_{4,e}; V) - I(V; U) = \frac{1}{2} \log_2 \left( 1 + (1 - \beta) P_2 \right).$$
Other mutual information terms can be obtained as follows:

\[ I(Y_{2,o}; U) = \frac{1}{2} \log_2 \left( 1 + \frac{a^2(1 - \alpha)P_1}{a^2\alpha P_1 + 1} \right); \]
\[ I(Y_{3,e}; U) = \frac{1}{2} \log_2 \left( 1 + \frac{P_1 + c^2\beta P_2 + 2c\sqrt{\beta P_2}P_1}{c^2(1 - \beta)P_2 + 1} \right); \]
\[ I(Y_{3,o}; X_o|U) = \frac{1}{2} \log_2 (1 + \alpha P_1). \]

Considering (B.2) and (B.3), the achievable rates for the PU and the SU can be obtained, and these rates are expressed in Theorem 7. Therefore, the proof is concluded.

\[ \Box \]

**B.5 Proof of Theorem 8**

In what follows, denote by \( \mathcal{E}_s(r) \) (or \( \mathcal{E}_s \) for brevity) the set of all eligible relays to participate as a DF relay for the source \( s \) when its spectral efficiency is set to \( r \). In other words, \( \mathcal{E}_s \) includes relays whose link budget to the source \( s \) allow an achievable rate greater than \( 2r \). The condition for the relay \( l \) to be in the set \( \mathcal{E}_s \) can be written as

\[ l \in \mathcal{E}_s(r) \leftrightarrow C \left( |h_{sl}|^2 \frac{P_s}{\sigma_N^2} \right) \geq 2r. \quad (B.7) \]

To derive the ASE, the probability of the event \( \{ R_s > r \} \) for a given \( r \) and a realization of the spatial relay distribution and shadowing needs to be obtained. To achieve this goal two different cases must be considered:
1. \( r \leq C(\gamma) \): In this case, the source transmits its message in one phase transmission without recruiting relays. Therefore, in this case, \( \Pr \{ c_{sd} > r \} = 1 \).

2. \( r > C(\gamma) \): In this case, a non-zero probability of \( \{ c_{sd} > r \} \) is feasible only if a spectral efficiency increment is achieved by employing relays in the two-phase transmission.

Considering the second case, by applying the law of total probability, it can be written

\[
\Pr \{ R_s > r \} = \sum_{k \geq 1} \Pr \{ R_s > r \mid |E_s| = k \} \Pr \{ |E_s| = k \}.
\]

Considering (5.10), we can write

\[
\Pr \{ R_s > r \mid |E_s(r)| = k \} = \Pr \{ c_s > 2r \mid |E_s(r)| = k \}
\]

where

\[
c_s = C \left( |h_{sd}|^2 \frac{P_s}{\sigma_N^2} + \left| \sum_{i \in \{s\} \cup E_s} h_{id} w_i \sqrt{P_i} \right|^2 / \sigma_N^2 \right)
\]

\[
= C \left( \gamma + \gamma \left| \sum_{i \in \{s\} \cup E_s} e^{j\hat{\theta}_i} \right|^2 \right).
\]

\( \hat{\theta}_i = \theta_i + \theta_{w_i} \) is the phase of the channel gain of UE \( i \) to the AP when its phase is shifted by the preceding weight \( w_i \). It can be seen form (B.9) that \( c_s \) is maximized when \( \left| \sum_{i \in \{s\} \cup E_s} e^{j\hat{\theta}_i} \right| \) is maximized. The maximum value of the latter quantity is \( k + 1 \) when all \( \hat{\theta}_i \)'s are equal for \( i \in \{s\} \cup E_s \). Given this assumption, we can write

\[
\Pr \{ c_s > 2r \mid |E_s(r)| = k \} \leq \Pr \{ C \left( \gamma [1 + (k + 1)^2] \right) > 2r \mid |E_s(r)| = k \}
\]

\[
= \Pr \{ k \geq k_r \mid |E_s(r)| = k \}
\]

(B.10)
where \( k_r \) can be easily computed and is given in the Theorem 8. Moreover, the event \(|\mathcal{E}_s(r)| = k\) can be translated as the event such that \( k \) UEs (indexed by \( l, l \in \{1, \cdots, N_{\text{UE}}\} \setminus \{s\} \)) are spatially distributed such that

\[
C\left( \frac{G}{d_{sl}^\alpha} 10^{\sigma_{dB} V_i/10} \frac{P_s}{\sigma_N^2} \right) \geq 2r \tag{B.11}
\]

where \( V_i \) is a normal Gaussian random variable. The inequality in (B.11) can be rearranged as

\[
d_{sl} e^{-\sigma V_i} \leq d_r \tag{B.12}
\]

where \( \sigma = 0.1 \ln(10) \sigma_{dB}/\alpha \), and

\[
d_r = \left[ \frac{GP_s/\sigma_N^2}{2^{2r} - 1} \right]^{1/\alpha}.
\]

So, the eligibility condition in (B.7) can be simplified as

\[
l \in \mathcal{E}_s(r) \leftrightarrow d_{sl} e^{-\sigma V_i} \leq d_r. \tag{B.13}
\]

For \( \sigma_{dB} = 0 \) (no shadowing), let \( \Psi_s(r) \) be the area of the disc centred at the location of the source \( s \) with radius \( d_r \) such that (B.11) holds (\( \Psi_s(r) \) defined in the Theorem 8). The probability of \( k \) Poisson arrivals in the area \( \Psi_s(r) \) can be obtained as

\[
\Pr \{ |\mathcal{E}_s(r)| = k \} = \frac{[\lambda \Psi_s(r)]^k}{k!} e^{-\lambda \Psi_s(r)}. \tag{B.14}
\]

The probability in (B.10) is either zero when \( k < k_r \) or one when \( k \geq k_r \). Therefore, the
complementary CDF of the rate $R_s$ can be upper bounded as

$$\Pr \{ R_s > r \} \leq \sum_{k \geq k_r} \Pr \{ |E_s(r)| = k \} = \sum_{k \geq k_r} \frac{[\lambda \Psi_s(r)]^k}{k!} e^{-\lambda \Psi_s(r)}. \quad (B.15)$$

Since $R_s$ is a non-negative random variable, we can express the expected value of $R_s$ as

$$E[R_s] = \int_0^{\infty} \Pr \{ R_s > r \} \, dr$$

$$= \int_0^{C(\gamma)} dr + \int_{C(\gamma)}^{\infty} \Pr \{ R_s > r \} \, dr$$

$$\leq C(\gamma) + \int_{C(\gamma)}^{\infty} \sum_{k \geq k_r} \frac{[\lambda \Psi_s(r)]^k}{k!} e^{-\lambda \Psi_s(r)} \, dr. \quad (B.16)$$

\[ \square \]

### B.6 Proof of Lemma 1

For the case with the log-normal shadowing ($\sigma_{dB} > 0$), we adopt an asymptotic method to find $\Pr \{ |E_s| = k \}$ in (B.14).

Assume that relays are randomly located according to a shrinking Bernoulli process on a disk with radios $d_{\text{max}}$ centred at the source. For this process, $2\pi \lambda \delta d_i$ is the probability that there exists one user in the distance interval of $[d_i - \delta, d_i)$ of the source node where $d_i = i\delta$. This process asymptotically approaches a Poisson process with rate $\lambda$ as $\delta \to 0 \quad [118]$. Let $n_{\text{max}} = \lfloor d_{\text{max}}/\delta \rfloor$. For $i = 1, \ldots, n_{\text{max}}$, let

$$E_i \triangleq \{ \text{exists one relay } l \text{ s.t. } d_{sl} e^{-\sigma V_l} \in [\delta (i - 1), \delta i] \} ,$$

$$E'_i \triangleq \{ \text{exists one relay at } d_i \} .$$
Furthermore, by definition, let

\[ p_{j \to i} = \Pr \{ j\delta e^{-\sigma V} \in [(i-1)\delta, i\delta) | E'_j] \} \] (B.17)

for \( i, j \in \{1, 2, \cdots, n_{\text{max}}\} \) where \( V \) is a normal Gaussian random variable. The probability \( p_{j \to i} \) can be intuitively interpreted as the probability that while a user is spatially located at distance \( d_j \) of the source, it is effectively within the distance \( [d_i - \delta, d_i) \) of the source when it is affected by the shadowing. This probability can be computed as

\[
p_{j \to i} = \Pr \{ (i-1)\delta \leq j\delta e^{\sigma V} < i\delta \} = \Pr \left\{ \frac{1}{\sigma} \ln \frac{i-1}{j} \leq V < \frac{1}{\sigma} \ln \frac{i}{j} \right\} = \frac{1}{2} \text{erf} \left( \frac{1}{\sqrt{2}\sigma} \ln \frac{i}{j} \right) - \frac{1}{2} \text{erf} \left( \frac{1}{\sqrt{2}\sigma} \ln \frac{i-1}{j} \right) \] (B.18)

where \((a)\) holds since \( V \) and \(-V\) have the same distributions. The probability of the event \( E_i \), \( p_i \), can be written as

\[
p_i = \sum_{j=1}^{n_{\text{max}}} \Pr \{ j\delta e^{-\sigma V} \in [(i-1)\delta, i\delta) | E'_j] \} \Pr \{ E'_j \} = 2\pi \lambda \delta^2 \sum_{j=1}^{\lfloor \delta_{\text{max}}/\delta \rfloor} j p_{j \to i}. \] (B.19)

Now, the event \( \{|E_s(r)| = k\} \) is tantamount to the event of \( k \) successes out of \( n \) YES/NO trials with probability of successes \( p_1, p_2, \cdots, p_n \) which has a Poisson binomial distribution \( (n = \lfloor d_r/\delta \rfloor) \). Let

\[
\Pr \{ |E_s(r)| = k \} = \lim_{\delta \to 0} \pi_k (r, \delta).
\]
Using a recursive method, we have

\[ \pi_k(r, \delta) = \begin{cases} 
\prod_{i=1}^{n} (1 - p_i), & k = 0, \\
\frac{1}{k} \sum_{j=1}^{k} (-1)^{j-1} \pi_{k-j}(r, \delta) T_j, & k > 0 
\end{cases} \]  

(B.20)

where

\[ T_j = \sum_{i=1}^{n} \left( \frac{p_i}{1 - p_i} \right)^j. \]

\[ \square \]

**B.7 Proof of Proposition**

We first claim that in Algorithm, \( r_{\text{min}} \) is a non-increasing function of \( |A_s| \). This claim can be verified as follows: considering that the set of available UEs for assistance, \( E_{\text{sorted}} \), is sorted with respect to the link rates \( R_{sl}, l \in E_{\text{sorted}} \), each addition to the \( A_s \) either limits \( r_{\text{min}} \) to a lower value or leaves it intact.

Moreover, we claim that in Algorithm, \( \gamma_{\text{AP}} \) is a non-decreasing function of \( |A_s| \) if the precoding resolution is high enough (\( N_w = \infty \)), and this claim can be easily verified by observing that in every step of the for loop, \( \gamma_{\text{AP}} \) either remains the same or becomes larger.

Considering these two claims, since \( r_{\text{min}} \) is increasing through the loop and \( \gamma_{\text{AP}} \) is decreasing through the loop, there is an optimal point in the loop that adding or deleting any UE to or from \( A_s \) would only decrease \( r_{\text{new}} \). \[ \square \]
B.8 Proof of Theorem 9

Considering the channel model in (5.17), the path loss in dB can be written as

\[ PL = k_0 + k_1 \ln d + \sigma_{dB} V \]

where \( k_0 = -10 \log_{10} G \), and \( k_1 = 10\alpha/\ln 10 \). Also, let the threshold loss \( l_{th} \) be defined as

\[ l_{th} = 10 \log_{10} \left( \frac{P_s}{\sigma^2_N} \left( 2^{2C(\gamma)} - 1 \right)^{-1} \right). \] (B.21)

To prove the main theorem, the following two lemmas are necessary:

**Lemma 2** The average number of eligible relays, \( E[|E_s|] \), is given as

\[ E[|E_s|] = \pi \lambda \exp \left( 2 \frac{l_{th} - k_0}{k_1} + 2 \frac{2C(\gamma)}{k_1^2} \right). \] (B.22)

**Proof:** See [120, Section III].

**Lemma 3** \( E[|E_s|] = \Theta(1) \).

**Proof:** From Lemma 2 it can be easily shown that \( E[|E_s|] \) remains bounded as \( N_{UE} \) grows.

Now, the main theorem can be proved. Let the total run time of the algorithm be divided into two following run-times: the run-time of the discovery of candidate relays (Algorithm 1 Line 2) denoted by \( T_{\text{disc.}}(N_{UE}) \) and the run-time of the rest of the algorithm excluding the precoding step denoted by \( T_{\text{rest}}(N_{UE}) \). Therefore, the average run-time can be written as

\[ E[T(N_{UE})] = E[T_{\text{disc.}}(N_{UE})] + E[T_{\text{rest}}(N_{UE})]. \] (B.23)

For the discovery step, all the relays has to be processed and their source-relay channel gains are compared to the threshold \( 2C(\gamma) \); therefore, this step is performed in \( \Theta(N_{UE}) \) steps.
The average run-time of the rest of the algorithm can be written as:

\[
E[T_{\text{rest}}(N_{UE})] = \Pr \left\{ |E_s| < \sqrt{N_{UE}} \right\} \mathcal{O} \left( \sqrt{N_{UE}} \log \sqrt{N_{UE}} \right) + \Pr \left\{ |E_s| > \sqrt{N_{UE}} \right\} \mathcal{O} \left( N_{UE} \log N_{UE} \right).
\]  \hfill (B.24)

Using the Markov inequality, it can be shown that

\[
\Pr \left\{ |E_s| > \sqrt{N_{UE}} \right\} \leq \frac{E[|E_s|]}{\sqrt{N_{UE}}}. \hfill (B.25)
\]

Therefore,

\[
E[T_{\text{rest}}(N_{UE})] = \mathcal{O} \left( \sqrt{N_{UE}} \log \sqrt{N_{UE}} \right) + \frac{E[|E_s|]}{\sqrt{N_{UE}}} \mathcal{O} \left( N_{UE} \log N_{UE} \right)
\]
\[
\overset{(a)}{=} \mathcal{O} \left( \sqrt{N_{UE}} \log N_{UE} \right) + \Theta(1) \mathcal{O} \left( N_{UE} \log N_{UE} \right)
\]
\[
= \mathcal{O} \left( \sqrt{N_{UE}} \log N_{UE} \right).
\]  \hfill (B.26)

where \((a)\) follows from Lemma 3. Consequently, the growth of the \(E[T_{\text{rest}}(N_{UE})]\) is slower than the \(E[T_{\text{disc.}}(N_{UE})]\) which makes the \(E[T_{\text{disc.}}(N_{UE})]\) the dominant term. Therefore, the overall complexity of the algorithm will be \(\Theta(N_{UE})\).

\[\square\]

### B.9  A Constructive Protocole to Achieve \(\tilde{r}_l\) in Definition 6.1.1

In this section we propose a protocol for message exchange between the UEs in a cluster and obtain the respective empirical intra-cluster rates \(\tilde{r}_{[\mathcal{I}]}\), \(l \in \mathcal{L}\). As mentioned, the message exchange between UEs is performed over \(\text{RAT}_2\) which is orthogonal to \(\text{RAT}_1\).

Let the total message exchange time \(T\) be divided into \(n_T\) time slots where

\[
n_T = \text{lcm} \left\{ N_l : l \in \mathcal{L} \right\}. \hfill (B.27)
\]
In each slot one UE per VD is transmitting, i.e., the total of |L| UEs are transmitting in each time slot. Each UE \( m \) occupies \( 1/|I_{l,m}| \) of slots. Given these transmissions, let \( \gamma_{ijt} \) be the received SINR of the UE \( i \) at UE \( j \) at time slot \( t \) considering all other transmitting UEs as noise. Since all the UEs in a VD are required to decode each other’s messages, the UE \( i \) rate is limited by the minimum SINR at the respective receivers. That is, the rate of the UE \( i \) at time slot \( t \) is

\[
 r_{it} = \min_{j \in I_{l,i}} \left\{ C(\gamma_{ijt}) \right\} . \tag{B.28}
\]

The effective rate of UE \( i \) can be obtained by averaging \( r_{it} \) over time. Note that for a time slot \( t \) in which UE \( i \) is not transmitting, \( r_{it} = 0 \). Therefore, we have

\[
 \bar{r}_{l} = \sum_{i \in I_{l}} \frac{1}{n_T} \sum_{t=1}^{n_T} r_{it} . \tag{B.29}
\]

### B.10 Proof of Proposition 2

The MSE of the VD \( l \) can be written as

\[
 \epsilon_{l} = E \left[ \| \hat{s}_{l} - s_{l} \|^{2} \right] 
\]

\[
 = E \left[ \text{Tr} \left( (\hat{s}_{l} - s_{l})(\hat{s}_{l} - s_{l})^{\dagger} \right) \right] 
\]

\[
 = g_{l}^{\dagger} \left( \sum_{l' \in L} \hat{H}_{k_{l}l'} \nu_{l'}^{\dagger} \nu_{l'}^{\dagger} \hat{H}_{k_{l}l'}^{\dagger} \right) g_{l} + I 
\]

\[
 - g_{l}^{\dagger} \hat{H}_{k_{l}l} \nu_{l} - \nu_{l}^{\dagger} \hat{H}_{k_{l}l}^{\dagger} g_{l} + \sigma_{N}^{2} g_{l}^{\dagger} g_{l} \tag{B.30}
\]

Considering the optimization problem in (6.1), the Lagrangian \( \mathcal{L} \) can be written as

\[
 \mathcal{L} (\{v_{m}\}, \{g_{l}\}, \{\lambda_{m}\}) = \sum_{l \in \mathcal{L}} w_{l} \epsilon_{l} + \sum_{m \in \mathcal{M}} \lambda_{m} (\|v_{m}\|^{2} - P_{m})
\]
where \( \{\lambda_m\} \) are Lagrangian multipliers. A (local) minimum of (6.11) must satisfy the KKT conditions [121]. By applying the rules for complex valued matrix derivation [122], it can be written

\[
\frac{\partial}{\partial \mathbf{v}^*_m} \mathcal{L} = \sum_{l \in \mathcal{L}} w_l \mathbf{g}_l \left( \sum_{l' \in \mathcal{L}, j \in \mathcal{I}_l'} \mathbf{h}_{k_{li} j} \mathbf{v}^*_l \mathbf{h}_{k_{lj}}^\dagger \right) \mathbf{g}_l \\
+ \frac{\partial}{\partial \mathbf{v}^*_m} \sum_{l \in \mathcal{L}} w_l \sum_{i \in \mathcal{I}_l} \mathbf{g}_l \mathbf{h}_{k_{li}} \mathbf{v}_i \\
+ \frac{\partial}{\partial \mathbf{v}^*_m} \sum_{l \in \mathcal{L}} w_l \sum_{i \in \mathcal{I}_l} \mathbf{v}_i \mathbf{h}_{k_{li}}^\dagger \mathbf{g}_l \\
+ \lambda_m \mathbf{v}_m. \tag{B.31}
\]

Therefore,

\[
\frac{\partial}{\partial \mathbf{v}^*_m} \mathcal{L} = \sum_{l \in \mathcal{L}} \mathbf{h}_{k_{lm}}^\dagger \mathbf{g}_l \sum_{i \in \mathcal{I}_{l,m} \setminus \{m\}} \mathbf{h}_{k_{im}} \mathbf{v}_m \\
+ \sum_{l \in \mathcal{L}} \sum_{i \in \mathcal{I}_{l,m}} \mathbf{h}_{k_{im}} \mathbf{g}_l \sum_{i \in \mathcal{I}_l \setminus \{m\}} \mathbf{h}_{k_{ii}} \mathbf{v}_i \\
+ \mathbf{h}_{a_m}^\dagger \mathbf{g}_a \mathbf{w}_a + \lambda_m \mathbf{v}_m. \tag{B.32}
\]

Applying \( \frac{\partial}{\partial \mathbf{v}^*_m} \mathcal{L} = 0 \), the precoding weights are obtained according to (6.12). Also, for \( \lambda_m > 0 \) we must have

\[
\lambda_m (\|\mathbf{v}_m\|^2 - P_m) = 0, \quad m \in \mathcal{M}; \\
\lambda_m \neq 0 \Rightarrow \mathbf{v}_m \mathbf{v}_m^\dagger - P_m = 0 \\
\Rightarrow \lambda_m^2 + \kappa_1 \lambda_m + \kappa_2 = 0; \tag{B.33}
\]

solving this second order equation for positive Lagrangian multipliers results in (6.16). Lastly, applying \( \frac{\partial}{\partial \mathbf{g}_l} \mathcal{L} = 0 \) results in the updates in (6.13).
B.11 Derivations of the Precoding Vectors with Per-Antenna Power Constraints

As mentioned, obtaining decoders and precoders that minimizes the weighted sum MSE is studied in [82, 83, 86, 89] for a sum power constraint. In this section, we find the decoder and precoder updates when there is a per-antenna power constraint. Our updates are similar to those in mentioned works except that the Lagrangians are obtained differently.

The optimization problem in (6.11) can be rearranged as

\[
\begin{align*}
\min & \quad \sum_{l \in \mathcal{L}} w_l \epsilon_l \\
\text{s.t.} & \quad \|v_l^\dagger e_i\|^2 \leq P_i, \quad i \in \mathcal{I}_l, l \in \mathcal{L}
\end{align*}
\]  
(B.33a)

(B.33b)

where \(e_i\) is a column vector with all its elements being zero except the \(i\)-th element being one. The Lagrangian of this optimization problem can be written as

\[
\mathcal{L}(\{v_l\}, \{g_l\}, \{\lambda_m\}) = \sum_{l \in \mathcal{L}} w_l \epsilon_l + \sum_{l \in \mathcal{L}, i \in \mathcal{I}_l} \lambda_l (v_l^\dagger e_i e_i^\dagger v_l - P_i)
\]

Following the same steps as those in the proof of Proposition 2, it can be shown that the following update for the precoding vector achieves a minimum of the optimization problem.

\[v_l = (\Xi_l + \Lambda_l)^{-1} \tilde{g}_l\]  
(B.34)

where

\[
\Xi_l = \sum_{l' \in \mathcal{L}} \tilde{H}_{k, l'}^\dagger g_{l'} w_{l'} g_{l'}^\dagger \tilde{H}_{k, l'},
\]  
(B.35)
\[ \Lambda_l = \sum_{i \in I_l} \lambda_i e_i e_i^\dagger, \quad (B.36) \]

and

\[ \tilde{g}_l = H_{k,l}^I g_l w_l. \quad (B.37) \]

The derivation of the decoding vectors \( g_l \) are similar, and it leads to the same updates as that in (6.13). Positive Lagrangian multipliers must satisfy

\[ \lambda_i \left( \| v_i^\dagger e_i \|^2 - P_i \right) = 0, \quad i \in I_l, l \in \mathcal{L}; \quad (B.38) \]

therefore,

\[ g_l^\dagger (\Xi_l + \Lambda_l)^{-1} e_i e_i^\dagger (\Xi_l + \Lambda_l)^{-1} g_l - P_i = 0 \quad (B.39) \]

for \( \lambda_m > 0 \). The set of non-linear equations in (B.39), can be solved using a quasi-newton solver. This equation might have multiple answers satisfying the positiveness of the Lagrangians. In this case of multiple choices, we adopt the first one with no particular ordering. If the system of non-linear equations in (B.39) can not be solved, \( \Lambda_l \leftarrow 0 \), and \( v_l \) is normalized to meet the power constraints.
Appendix C

Full-Duplex Causal Cognitive Radio

In this appendix, a two-user CCR channel is studied, in which the SU has causal knowledge about the message being sent by the PU. The SU can also cooperate with the PU in the FD mode. An inner bound on the capacity region of such channels is established by employing a coding strategy consisting of block Markov superposition and DPC and backward decoding. An illustrative example in the Gaussian case is provided.

C.1 Definitions and Channel Modelling

Consider the discrete memoryless CCR as illustrated in Fig. C.1. The CCR comprises 4 nodes where there is a causal unidirectional cooperation between node 1 and node 2. This channel is denoted by \((\mathcal{X}_1 \times \mathcal{X}_2, p(y_2, y_3, y_4|x_1, x_2), \mathcal{Y}_2 \times \mathcal{Y}_3 \times \mathcal{Y}_4)\), where \(\mathcal{X}_1\) and \(\mathcal{X}_2\) are the finite input alphabets of the PU and SU respectively, \(\mathcal{Y}_2, \mathcal{Y}_3, \text{ and } \mathcal{Y}_4\) are the finite output alphabets of nodes 2, 3 and 4 respectively, and \(p(\cdot, \cdot, \cdot|x_1, x_2)\) is a collection of probability distributions on \(\mathcal{Y}_2 \times \mathcal{Y}_3 \times \mathcal{Y}_4\). Following the standard notation adopted in \([114]\), we define a \((2^{nR_1}, 2^{nR_2}, n)\) code for the CCR in the following.

**Definition C.1.1** A \((2^{nR_1}, 2^{nR_2}, n)\) code for the CCR consists of two message sets \(\mathcal{W}_1 = \)

---

\(^1\)The results of this chapter are also presented in \([22]\).
\{1, 2, \ldots, 2^{nR_1}\} for the PU and \(\mathcal{W}_2 = \{1, 2, \ldots, 2^{nR_2}\}\) for the SU, an encoding function 
\(X_1 : \mathcal{W}_1 \mapsto \mathcal{X}_1^n\), a set of encoding functions \(\{f_i\}_{i=1}^n\) such that \(x_{2i} = f_i(w_2, y_2^{i-1})\), with \(w_2 \in \mathcal{W}_2\), and two decoding functions \(g_1 : \mathcal{Y}_3^n \mapsto \mathcal{W}_1\), \(g_2 : \mathcal{Y}_4^n \mapsto \mathcal{W}_2\).

The channel is assumed to be memoryless; therefore, for any choice of \(p(w_1), p(w_2)\) and encoding functions, the probability mass function over \(\mathcal{W}_1 \times \mathcal{W}_2 \times \mathcal{X}_1^n \times \mathcal{X}_2^n \times \mathcal{Y}_3^n \times \mathcal{Y}_4^n\) is given by

\[
p(w_1, w_2, x_1^n, x_2^n, y_2^n, y_3^n, y_4^n) = p(w_1)p(w_2) \times \prod_{i=1}^n p(x_{1i}|w_1)p(x_{2i}|w_2, y_2^{i-1})p(y_{2i}, y_{3i}, y_{4i}|x_{1i}, x_{2i}).
\]

The average probability of error of the code is defined as

\[
P_e^{(n)} = \frac{1}{2^{n(R_1+R_2)}} \sum_{w_1, w_2} \Pr\left\{\begin{array}{l}
g_1(Y_3^n) \neq w_1 \\
\cup g_2(Y_4^n) \neq w_2
end{array} \mid w_1, w_2, \text{sent}\right\}.
\]

The probability of error is obtained under the assumption that codewords \(w_1\) and \(w_2\) are uniformly distributed in \(\mathcal{W}_1\) and \(\mathcal{W}_2\), respectively.

**Definition C.1.2** A rate pair \((R_1, R_2)\) is called achievable for the [CCR] if there is a sequence of \((2^{nR_1}, 2^{nR_2}, n)\) codes with \(P_e^{(n)} \to 0\). The capacity region \(C\) of the [CCR] is the union of set of all achievable rates.

Nodes 1 and 2 are also referred to as the PU and SU respectively, as depicted in Fig. [C.1]

### C.2 Main Results

#### C.2.1 An Achievable Rate Region for the Causal CR

Let \(U_1, U_2, V_1, V_2, X_{11}, X_{12}, X_1, X_2,\) and \(Q\) be RVs defined over the finite alphabet sets \(\mathcal{U}_1, \mathcal{U}_2, \mathcal{V}_1, \mathcal{V}_2, \mathcal{X}_{11}, \mathcal{X}_{12}, \mathcal{X}_1, \mathcal{X}_2,\) and \(Q\) respectively where \(Q\) plays the role of time sharing RV.
Figure C.1: Channel model for a CCA, a causal configuration for the CR. $x_1$ and $x_2$ are the channel inputs, and $y_2$, $y_3$, $y_4$ are the channel outputs.

Let the collection of RVs be $Z = (Q, U_1, U_2, V_1, V_2, X_{11}, X_{12}, X_1, X_2, Y_2, Y_3, Y_4)$. Denote by $\mathcal{P}$ the set of all joint probability distributions $p(\cdot)$ on $Z$ that can be factored as

\[
p(z) = p(q)p(u_1|q)p(u_2|q)p(x_{11}|u_1, q)p(x_{12}|u_2, q) \\
p(x_1|x_{11}, x_{12}, u_1, u_2, q)p(v_1|q)p(v_2|q) \\
p(x_2|v_1, v_2, u_1, u_2, q)p(y_2, y_3, y_4|x_1, x_2).
\]  

(C.1)
Let $\mathcal{R}(Z)$ be the set of all positive rate pairs $(R_1, R_2)$ such that $R_1 = R_{11} + R_{12}$, $R_2 = R_{21} + R_{22}$, and

$$R_{11} \leq I(Y_2; U_1 X_{11} | U_2 X_{12} Q),$$  \hspace{1cm} (C.2)

$$R_{12} \leq I(Y_2; U_2 X_{12} | U_1 X_{11} Q),$$  \hspace{1cm} (C.3)

$$R_{11} + R_{12} \leq I(Y_2; U_1 X_{11} U_2 X_{12} | Q),$$  \hspace{1cm} (C.4)

$$R_{21} \leq L_1 - I(V_1; U_1 U_2 | Q),$$  \hspace{1cm} (C.5)

$$R_{22} \leq L_2 - I(V_2; U_1 U_2 | Q),$$  \hspace{1cm} (C.6)

$$R_{21} + R_{22} \leq L_1 + L_2 - [I(V_1; V_2 | Q) + I(V_1 V_2; U_1 U_2 | Q)],$$  \hspace{1cm} (C.7)

$$R_{11} \leq I(Y_3 V_1; U_1 X_{11} | U_2 X_{12} Q),$$  \hspace{1cm} (C.8)

$$R_{12} \leq I(Y_3 V_1; U_2 X_{12} | U_1 X_{11} Q),$$  \hspace{1cm} (C.9)

$$R_{11} + R_{12} \leq I(Y_3 V_1; U_1 X_{11} U_2 X_{12} | Q),$$  \hspace{1cm} (C.10)

$$R_{11} + L_1 \leq I(Y_3 U_2 X_{12}; V_1 U_1 X_{11} | Q),$$  \hspace{1cm} (C.11)

$$R_{12} + L_1 \leq I(Y_3 U_1 X_{11}; V_1 U_2 X_{12} | Q),$$  \hspace{1cm} (C.12)

$$R_{11} + R_{12} + L_1 \leq I(Y_3; V_1 U_1 X_{11} U_2 X_{12} | Q),$$  \hspace{1cm} (C.13)

$$L_1 \leq I(Y_4 V_2 U_2 X_{12}; V_1 | Q),$$  \hspace{1cm} (C.14)

$$L_2 \leq I(Y_4 V_1 U_2 X_{12}; V_2 | Q),$$  \hspace{1cm} (C.15)

$$R_{12} + L_1 \leq I(Y_4 V_2; U_2 X_{12} V_1 | Q),$$  \hspace{1cm} (C.16)

$$R_{12} + L_2 \leq I(Y_4 V_1; U_2 X_{12} V_2 | Q),$$  \hspace{1cm} (C.17)

$$L_1 + L_2 \leq I(Y_4 U_2 X_{12}; V_1 V_2 | Q),$$  \hspace{1cm} (C.18)

$$R_{12} + L_1 + L_2 \leq I(Y_4; U_2 X_{12} V_1 V_2 | Q).$$  \hspace{1cm} (C.19)

**Theorem 11** For any $p(\cdot) \in \mathcal{P}$, the region $\mathcal{R}(Z)$ is an achievable rate region for the CCR described in Section C.1, i.e., $\bigcup_{Z \sim p(\cdot)} \mathcal{R}(Z) \subseteq C$. 

Proof: See Section C.4

C.2.2 Relationship between Theorem 11 and Existing Results

The Relay Channel

When the message set of the second transmitter is empty, nodes 1, 2, and 3 in CCR can be reduced to a relay channel. The rate region in Theorem 11 can be reduced to the DF achievable rate for the relay channel [24]. By nullifying random variables $U_2, X_{12}, V_1, V_2$, and letting $X_1 = X_{12}, X_2 = U_1, R_{12} = 0$, we will have $R_1 \leq \min \{I(Y_3; X_1 X_2), I(Y_2; X_1 | X_2)\}$, which is the capacity of the relay channel given that the channel is degraded [24].

The Interference Channel

If we omit the channel output $Y_2$ from the channel model, the model is reduced to the conventional interference channel. The best inner bound on the capacity of the interference channel is given by Han and Kobayashi (HK) [25]. To achieve the HK rate region, we nullify RVs $U_1$ and $U_2$, and let $L_1 = R_{21}, L_2 = R_{22}$. By doing so, we can readily conclude that $V_1$ and $V_2$ are independent of $X_1$ and thus our rate region reduces to the HK region.

The Interference Channel with Conferencing

As mentioned in the introduction, CCR is analogous to the ICC since in both models causal knowledge about the message of the PU is available at the SU. An achievable rate region for the ICC is obtained in [5, Theorem 1]. However, this rate region is strictly suboptimal when the channel output $y_3$ is almost independent of $x_1$, i.e., the channel probability distribution $p(y_2, y_3, y_4 | x_1, x_2)$ can be factored as $p(y_2, y_4 | y_3, x_1, x_2)p(y_3 | x_2)$. This situation can, for instance, arise when nodes 1 and 3 are geometrically far from each other. In this scenario, the rates associated with private and common messages in [5, Theorem 1] are zero and the PU can convey its message to node 3 only via node 2. Had this cooperative message been split and
treated similar to the HK coding scheme, Theorem 1 in [5] would have potentially resulted in an improved achievable rate region.

Using Theorem 11, the achievable rate region for an ICC can be obtained as following. First, let the CCR be modeled as described in Section C.1. Apply Theorem 11 to compute the rate region of the CCR. Then, in the channel modeled in Section C.1 change the roles of node 1 and node 2. In other words, let node 1 be the SU which can overhear through the channel. Apply Theorem 11 with proper modifications to compute the achievable rate region. Finally, declare the convex hull of these two regions as an achievable rate region for the ICC.

As illustrated in Fig. C.2(b) the region achieved by this approach includes that of [5] under certain channel conditions.

C.2.3 On The MIMO Broadcast Channel

The capacity region of a Gaussian MIMO broadcast channel (GMIMO-BC) with two antennas at the transmitter and single receiver antenna in each of two receivers (as modeled in [124]) is given below [124, Theorem 5].

**Lemma 4** The capacity region of a GMIMO-BC is the convex hull of following set of rate pairs

\[
\left\{(R_1, R_2) : 0 \leq R_1 \leq \frac{1}{2} \log \left( \frac{|H_1B_1H_1^T + N_1|}{|N_1|} \right), \right. \\
0 \leq R_2 \leq \frac{1}{2} \log \left( \frac{|H_2(B_1 + B_2)H_2^T + N_2|}{|H_2B_1H_2^T + N_2|} \right) \right\} \bigcup \\
\left\{(R_1, R_2) : 0 \leq R_1 \leq \frac{1}{2} \log \left( \frac{|H_1(B_1 + B_2)H_1^T + N_1|}{|H_1B_2H_1^T + N_1|} \right), \right. \\
0 \leq R_2 \leq \frac{1}{2} \log \left( \frac{|H_2B_2H_2^T + N_2|}{|N_2|} \right) \right\}
\]

for all positive semidefinite matrices $B_1$ and $B_2$ satisfying $B_1 + B_2 \preceq S$.

In this lemma, $S = \begin{bmatrix} P_1 & g \\ g & P_2 \end{bmatrix}$ is a positive semidefinite matrix indicating transmit antennas covariance matrix. $P_1$ and $P_2$ are individual (per antenna) power constraints and $g$ is an
arbitrary constant satisfying $g^2 \leq P_1 P_2$. $H_1 = \begin{bmatrix} h_{31} & h_{32} \end{bmatrix}$ and $H_2 = \begin{bmatrix} h_{41} & h_{42} \end{bmatrix}$ are the channel gain matrices.

Accordingly, two antennas at the transmitter of the MIMO-BC can be treated as nodes 1 and 2 in the CCR respectively, and two receivers can respectively be treated as nodes 3 and 4 in the CCR. Obviously, the MIMO-BC achieves a larger capacity region than that of the CCR since in the MIMO-BC, full non-causal knowledge about all the messages being transmitted is available to each antenna (node). Hence, under the same channel conditions, the capacity region of the MIMO-BC is an upper bound on the capacity region of the CCR. We use this fact to obtain the upper bound shown in Fig. C.2(b).

### C.3 The Gaussian Channel

In a Gaussian CCR, the input-output relationship can be mathematically expressed as

\[
Y_2 = h_{21}X_1 + Z_2, \quad (C.20)
\]
\[
Y_3 = h_{31}X_1 + h_{32}X_2 + Z_3, \quad (C.21)
\]
\[
Y_4 = h_{41}X_1 + h_{42}X_2 + Z_4 \quad (C.22)
\]

where $Y_2$, $Y_3$ and $Y_4$ are channel outputs. Furthermore, $X_1$ and $X_2$ are zero mean Gaussian channel inputs subject to power constraints: $E[|X_1|^2] \leq P_1$ and $E[|X_2|^2] \leq P_2$ respectively. The noise powers are subject to $E[|Z_2|^2] \leq N_2$, $E[|Z_3|^2] \leq N_3$, and $E[|Z_4|^2] \leq N_4$. The channel coefficients $h_{31}$, $h_{32}$, $h_{41}$, and $h_{42}$ are assumed to be known.

The generated codebook can be mapped into the Gaussian RVs as

\[
X_1 = \sqrt{\alpha_1 P_1} U_1 + \sqrt{\alpha_2 P_1} X'_{11} + \sqrt{\alpha_3 P_1} U_2 + \sqrt{\alpha_4 P_1} X'_{12}, \quad (C.23)
\]
\[
X_2 = \sqrt{\beta_1 P_2} X'_{21} + \sqrt{\beta_2 P_1} X'_{22} + \sqrt{\beta_3 P_1} U_1 + \sqrt{\beta_4 P_1} U_2 \quad (C.24)
\]

where $\sum_{i=1}^{4} \alpha_i = 1$, $\alpha_i \geq 0$, $\sum_{i=1}^{4} \beta_i = 1$, $\beta_i \geq 0$ and $U_1, U_2, X'_{11}, X'_{12}, X'_{21}, X'_{22}$ are Gaussian.
RVs with mean zero and variance unit. With the mapping defined in (C.23) and (C.24), the channel outputs $Y_3$ and $Y_4$ can be rewritten as

\begin{align*}
Y_3 &= \sqrt{l_1} U_1 + \sqrt{l_2} X_{11}' + \sqrt{l_3} U_2 + \sqrt{l_4} X_{12}' \\
&\quad + \sqrt{l_5} X_{21}' + \sqrt{l_6} X_{22}' + Z_3 \\
Y_4 &= \sqrt{f_1} U_1 + \sqrt{f_2} X_{11}' + \sqrt{f_3} U_2 + \sqrt{f_4} X_{12}' \\
&\quad + \sqrt{f_5} X_{21}' + \sqrt{f_6} X_{22}' + Z_4
\end{align*}

where $l_i, f_i$ can be determined by substituting (C.23), (C.24) into (C.21), and (C.22). The RVs $V_1$ and $V_2$ can be mapped as

\begin{align*}
V_1 &= \sqrt{\beta_1} P_2 X_{21}' + \gamma_1 S_1, \quad V_2 = \sqrt{\beta_2} P_2 X_{22}' + \gamma_2 S_2.
\end{align*}

The best values for $\gamma_1, \gamma_2, S_1,$ and $S_2$ can be found by considering a performance cost criterion (such as the sum rate) and solving the corresponding optimizing problem. This method is not, however, tractable, and for the sake of illustration, the following lemma is used to find particular values for $\gamma_1, \gamma_2, S_1,$ and $S_2$. This lemma is a slightly modified version of Costa’s DPC results [29].

**Lemma 5** *In a Gaussian memoryless point to point channel with state $S^n$ known at the transmitter with channel output $Y = aX + S + Z$ where $E[X^2] \leq P, E[S^2] \leq Q, E[Z^2] \leq N$, the choice of $V = X + \gamma S$ with $\gamma = aP/(a^2P + N)$ achieves the point to point capacity $0.5\log(1 + a^2P/N)$ as if there is no state present.*

Now, a Costa successive pre-coding (Lemma 5) can be applied to $V_1$ and $V_2$. While encoding $V_1$, the codewords $U_1, U_2$ are as known states. Therefore, the known state is $S_1 = \sqrt{f_1} U_1 + \sqrt{f_2} U_2$, and $\gamma_1$ can be written as

\begin{align*}
\gamma_1 &= \frac{h_{42} \beta_1 P_2}{h_{42} \beta_1 P_2 + \underbrace{N_4 + f_2 + f_4 + f_6}}.
\end{align*}

associated with noise power
Next, Lemma 5 is applied to $V_2$ with the known state $S_2 = \sqrt{f_1}U_1 + \sqrt{f_3}U_2 + \sqrt{f_5}X_{21}'$, and therefore, $\gamma_2$ can be written as

$$\gamma_2 = \frac{h_{42}\beta_2 P_2}{h_{42}^2\beta_2 P_2 + N_4 + f_2 + f_4}.$$ 

Fig. C.2(a) demonstrates a comparison between the results on CCR in [125] and Theorem 11. As can be seen, the new coding method benefiting from DPC in Theorem 11 resulted in a larger achievable rate region including the one in [125]. Fig. C.2(b) illustrates a comparison between the capacity region of GMIMO-BC, an achievable rate region for the ICC in [5], an achievable rate region in Theorem 11 for the CCR as discussed in Section C.2.2, and the HK region for the IC. As can be observed, the rate region in Theorem 11 includes that in [5]. The capacity region of the GMIMO-BC can be thought as an upper bound on the capacity region of the CCR.

### C.4 Proof of Theorem 11

A block-Markov encoding technique is used at the transmitters; messages are transmitted in two consecutive blocks. Also, each message set $\mathcal{W}_1$ and $\mathcal{W}_2$ is split into two parts. We assume that the message of sender 1 in the previous block is known to the sender 2 (see the decoding section). This known information can be considered as known state to the second sender. The binning method is used to mitigate the effect of this state at node 4. Both receivers decode the message by applying backward decoding. A detailed proof of Theorem 11 is given as follows.

#### C.4.1 Codebook Generation

At the primary sender, the rate splitting and superposition methods are used; therefore, the message $w_1$ is divided into two independent parts $w_{11}$ and $w_{12}$. The codebook is generated as following.
- Generate $2^{nR_1}$ codewords $u_1^n \overset{\text{i.i.d.}}{\sim} \prod_{i=1}^{n} p(u_{1,i})$ and index them as $u_1^n(w_{11})$ where $w_{11} \in \{1, \ldots, 2^{nR_1}\}$.

- For each message $w_{11}$, generate $2^{nR_1}$ codewords $x_{11}^n \overset{\text{i.i.d.}}{\sim} \prod_{i=1}^{n} p(x_{11,i}|u_{1,i}(w_{11}))$ and index them as $x_{11}^n(w_{11}^{'}, w_{11})$ where $w_{11}^{' \in} \{1, \ldots, 2^{nR_1}\}$.

- Generate $2^{nR_2}$ codewords $u_2^n \overset{\text{i.i.d.}}{\sim} \prod_{i=1}^{n} p(u_{2,i})$ and index them as $u_2^n(w_{12})$ where $w_{12} \in \{1, \ldots, 2^{nR_2}\}$.

- For each message $w_{12}$, generate $2^{nR_2}$ codewords $x_{12}^n \overset{\text{i.i.d.}}{\sim} \prod_{i=1}^{n} p(x_{12,i}|u_{2,i}(w_{12}))$ and index them as $x_{12}^n(w_{12}^{'}, w_{12})$ where $w_{12}^{'} \in \{1, \ldots, 2^{nR_2}\}$.

- For each message tuple $(w_{11}^{'}, w_{12}^{'}, w_{11}, w_{12})$, generate a codeword $x_1^n \overset{\text{i.i.d.}}{\sim} \prod_{i=1}^{n} p(x_{1,i}|x_{11,i}(w_{11}^{'}, w_{11}), x_{12,i}(w_{12}^{'}, w_{12}))$ and index it as $x_1^n(w_{11}^{'}, w_{12}^{'}, w_{11}, w_{12})$.

- Generate $2^{nL_1}$ codewords $v_1^n \overset{\text{i.i.d.}}{\sim} \prod_{i=1}^{n} p(v_{1,i})$ and index them as $v_1^n(l_1)$ where $l_1 \in \{1, \ldots, 2^{nL_1}\}$ and $L_1 \geq R_{21}$.

- Generate $2^{nL_2}$ codewords $v_2^n \overset{\text{i.i.d.}}{\sim} \prod_{i=1}^{n} p(v_{2,i})$ and index them as $v_2^n(l_2)$ where $l_2 \in \{1, \ldots, 2^{nL_2}\}$ and $L_2 \geq R_{22}$.

- Construct $2^{n(R_{21}+R_{22})}$ bins and index them as $(w_{21}, w_{22})$ where $w_{21} \in \{1, \ldots, 2^{nR_{21}}\}$ and $w_{22} \in \{1, \ldots, 2^{nR_{22}}\}$. Randomly and identically, distribute $2^{n(L_1+L_2)}$ codewords $\{v_1^n(l_1), v_2^n(l_2)\}$ into $2^{n(R_{21}+R_{22})}$ constructed bins. Denote by $B(w_{21}, w_{22})$ the set of all the message index pairs $(l_1, l_2)$ such that $\{v_1^n(l_1), v_2^n(l_2)\}$ belongs to bin with index $(w_{21}, w_{22})$. Also, let $B(w_{21})$ be the set of all message indices $l_1$ such that codeword $v_1^n(l_1)$ belongs to bin $(w_{21}, w_{22})$ where $w_{22} \in \{1, \ldots, 2^{nR_{22}}\}$. In a similar manner, let $B(w_{22})$ be the set of all message indices $l_2$ such that codeword $v_2^n(l_2)$ belongs to bin $(w_{21}, w_{22})$ where $w_{21} \in \{1, \ldots, 2^{nR_{21}}\}$.

- For each message tuple $(l_1, l_2, w_{11}, w_{12})$, generate an i.i.d. codeword $x_2^n$ according to probability distribution $\prod_{i=1}^{n} p(x_{2,i}|v_{1,i}(l_1), v_{2,i}(l_2), u_{1,i}(w_{11}), u_{2,i}(w_{12}))$ and index it as
Lastly, the generated codebook is revealed to all transmitters and receivers.

C.4.2 Encoding

The encoder at node 2 is required to be able to perfectly decode the message transmitted by node 1 (see the decoding section). At node 1, a block Markov superposition coding method is used, and transmission is done in $B$ blocks. In the first block, node 1 transmits codeword $x_1^n(w_{11}, w_{12}, \emptyset, \emptyset)$. In block $b$, node 1 transmits $x_1^n(w_{11,b}, w_{12,b}, w_{11,b-1}, w_{12,b-1})$. At the last block, it transmits a previously known message that is $(w_{11,B}, w_{12,B}) = (1, 1)$, and therefore, the codeword $x_1^n(w_{11,B-1}, w_{12,B-1}, 1, 1)$ is transmitted.

In order to transmit the message pair $(w_{21,b}, w_{22,b})$ in block $b$, the encoder at node 2 seeks in bin $B(w_{21,b}, w_{22,b})$ to find indices $(l_{1,b}, l_{2,b})$ whose codewords $(v_1^n(l_{1,b}), v_2^n(l_{2,b}))$ are jointly typical with $(u_1^n(w_{11,b-1}), u_2^n(w_{12,b-1}))$. If no such codewords can be found, an error is declared. For notational convenience, let $u_1^n, u_2^n$ denote $u_1^n(w_{11,b-1}), u_2^n(w_{12,b-1})$ respectively, and $B(\cdot, \cdot)$ denote $B(w_{21,b}, w_{22,b})$. An error is declared if any of the following events happens:

\begin{align}
\{ \forall l_1 \in B(w_{21,b}) : (v_1^n(l_1), u_1^n, u_2^n) \in T_\epsilon(n) \}, \quad \text{(C.25)} \\
\{ \forall l_2 \in B(w_{22,b}) : (v_2^n(l_2), u_1^n, u_2^n) \in T_\epsilon(n) \}, \quad \text{(C.26)} \\
\{ \forall (l_1, l_2) \in B(\cdot, \cdot) : (v_1^n(l_1), v_2^n(l_2), u_1^n, u_2^n) \in T_\epsilon(n) \}. \quad \text{(C.27)}
\end{align}

In the last block, a previously known message pair $(w_{21,B}, w_{22,B}) = (1, 1)$ is sent.
C.4.3 Decoding

At node 2, it is declared that $\tilde{w}_{11,b}, \tilde{w}_{12,b}$ is sent if there is a unique message pair $(\tilde{w}_{11,b}, \tilde{w}_{12,b})$ such that

$$
(u^n_1(w_{11,b-1}), x^n_{11}(\tilde{w}_{11,b}, w_{11,b-1}),
 u^n_2(w_{12,b-1}), x^n_{12}(\tilde{w}_{12,b}, \hat{w}_{12,b-1}), Y^n_{2,b}) \in T^{(n)}_e; \quad (C.28)
$$

otherwise, an error is declared. However, the probability of error can be shown to be vanished as $n \to \infty$ if (C.2)–(C.4) hold.

At node 3, the backward decoding is used. The encoder at this node starts decoding after all $B$ blocks of data are received. The encoder declares the message tuple $(\hat{w}_{11,b-1}, \hat{w}_{12,b-1}, \hat{l}_{1,b})$ is sent if there is only one such a message tuple such that

$$
(v^n_1(\hat{l}_{1,b}), v^n_1(\hat{l}_{1,b}), Y^n_{3,b}) \in T^{(n)}_e; \quad (C.29)
$$

otherwise, an error is declared. Note that the backward decoding ensures that in above mentioned decoding phase the message pair $(w_{11,b}, w_{12,b})$ is known.

Similarly, the backward decoding is used at node 4. This node starts decoding after all $B$ blocks of data are received. The encoder declares the message tuple $(\hat{w}_{12,b-1}, \hat{l}_{1,b}, \hat{l}_{2,b})$ is sent if there is only one such a message tuple such that

$$
(v^n_1(\hat{l}_{1,b}), v^n_2(\hat{l}_{2,b}), Y^n_{4,b}) \in T^{(n)}_e; \quad (C.30)
$$

otherwise, an error is declared. Again the backward decoding ensures that in the above men-
tioned decoding phase the message \( w_{12,b} \) is known.

### C.4.4 Error Probability Analysis

It can be shown that the probability of events in Eqs. (C.25)–(C.27) are respectively less than or equal to the following quantities

\[
(1 - 2^{-n(I(V_1;U_1,U_2) - \epsilon)})^{\mathcal{B}(w_{21})},
\]

\[
(1 - 2^{-n(I(V_2;U_1,U_2) - \epsilon)})^{\mathcal{B}(w_{22})},
\]

and

\[
(1 - 2^{-n(I(V_1;U_1,U_2) - \epsilon)})2^{-n(I(V_2;V_1,U_1,U_2) - \epsilon)})^{\mathcal{B}(w_{21},w_{22})}.
\]

Given that the size of the bins \( \mathcal{B}(w_{21}), \mathcal{B}(w_{22}), \) and \( \mathcal{B}(\cdot,\cdot) \) are respectively equal to \( 2^{n(L_1 - R_{21})}, 2^{n(L_2 - R_{22})}, \) and \( 2^{n(L_1 + L_2 - R_{21} - R_{22})} \), the probability of occurring an error at node 2 can be arbitrary small as \( n \to \infty \) if (C.5)–(C.7) hold. Furthermore, it can be shown from (C.29) that the probability of happening an error at node 3 can be arbitrary small as \( n \to \infty \) if (C.8)–(C.13) hold. Furthermore, by considering Eq. (C.30), it can be shown that the probability of happening an error at node 4 can be arbitrary small as \( n \to \infty \) if (C.14)–(C.19) hold. \( \square \)
Figure C.2: (a) Comparison between the achievable rate regions of (i) the CCR in [125] and (iii) Theorem 11 when $P_1 = 6$, $P_2 = 1.5$, and the achievable rate regions of (ii) the CCR in [125] and (iv) Theorem 11 when $P_1 = 1.5$, $P_2 = 6$. The channel parameters are $h_{31} = h_{42} = 1$, $h_{32} = h_{41} = 0.74$, $h_{21} = 4$, and $N_2 = N_3 = N_4 = 1$. (b) Comparison between (i) the capacity region of the GMIMO-BC [124] in Section C.2.3, (ii) ICC achievable rate region in [5], (iii) HK rate region, (iv) achievable rate region in Theorem 11 when only transmitter 2 is cognitive, (v) achievable rate region in Theorem 11 when only transmitter 1 is cognitive (see Section C.2.2), (vi) convex hull of (iv) and (v). The channel parameters are $h_{31} = h_{42} = 1$, $h_{32} = h_{41} = 0.74$, $h_{21} = h_{12} = 4$, $P_1 = 6$, $P_2 = 1.5$, and $N_2 = N_3 = N_4 = 1$. 
Bibliography


[26] K. Marton, “A coding theorem for the discrete memoryless broadcast channel,”


