Enhancing Sensing and Channel Access in Cognitive Radio Networks

Thesis by
Doha Raik Hamza Mohamed

In Partial Fulfillment of the Requirements
For the Degree of
Doctor of Philosophy
(Electrical Engineering)

King Abdullah University of Science and Technology, Thuwal,
Kingdom of Saudi Arabia

May, 2014
The thesis of Doha Raik Hamza Mohamed is approved by the examination committee

Committee Chairperson: Prof. Mohamed-Slim Alouini
Committee Member: Prof. Khaled Salama
Committee Member: Prof. Raul Tempone
Committee Member: Prof. Karim Abed-Meraim
ABSTRACT

Enhancing Sensing and Channel Access in Cognitive Radio Networks

Doha Raik Hamza Mohamed

Cognitive radio technology is a promising technology to solve the wireless spectrum scarcity problem by intelligently allowing secondary, or unlicensed, users access to the primary, licensed, users’ frequency bands. Cognitive technology involves two main tasks: 1) sensing the wireless medium to assess the presence of the primary users and 2) designing secondary spectrum access techniques that maximize the secondary users’ benefits while maintaining the primary users’ privileged status. On the spectrum sensing side, we make two contributions. First, we maximize a utility function representing the secondary throughput while constraining the collision probability with the primary below a certain value. We optimize therein the channel sensing time, the sensing decision threshold, the channel probing time, together with the channel sensing order for wideband primary channels. Second, we design a cooperative spectrum sensing technique termed sensing with equal gain combining whereby cognitive radios simultaneously transmit their sensing results to the fusion center over multipath fading reporting channels. The proposed scheme is shown to outperform orthogonal reporting systems in terms of achievable secondary throughput and to be robust against phase and synchronization errors. On the spectrum access side, we make four contributions. First, we design a secondary scheduling scheme with the
goal of minimizing the secondary queueing delay under constraints on the average secondary transmit power and the maximum tolerable primary outage probability. Second, we design another secondary scheduling scheme based on the spectrum sensing results and the primary automatic repeat request feedback. The optimal medium access probabilities are obtained via maximizing the secondary throughput subject to constraints that guarantee quality of service parameters for the primary. Third, we propose a three-message superposition coding scheme to maximize the secondary throughput without degrading the primary rate. Cognitive relaying is employed as an incentive for the primary network. The scheme is shown to outperform a number of reference schemes such as best relay selection. Finally, we consider a network of multiple primary and secondary users. We propose a three-stage distributed matching algorithm to pair the network users. The algorithm is shown to perform close to an optimal central controller, albeit at a reduced computational complexity.
ACKNOWLEDGEMENTS

I thank Allah (SWT) for enabling me to finish this thesis.

I thank my first advisor, Prof Sonia Aïssa for giving me the opportunity to be at KAUST and for her enduring support and care. I also sincerely thank my current advisor Prof Mohamed-Slim Alouini for his generous help, guidance, and dedication. I thank my research proposal committee members Professors Basem Shihada and Tareq AlNaffouri for their helpful feedback and suggestions.

I am particularly thankful for my colleagues Amal Hyadi, Emna Zedini, Hessa AlQuwaiee, Fatma Benkhelifa, Hina Tabassum, Fahd Khan, Ammar Zafar, Laila Affy, Mohammad Khafagy and Imran Shafique Ansari for engaging technical discussions and for being always very quick to the help of a helpless PhD student. I thank Dr. Ki-Hong Park for his research collaboration. I learned a lot from him.

Throughout my life, I have been blessed with the knowledge and friendship of many wonderful people. Thank you Amal, Hend, Sara, Sawsan, Imene, Sally, Samar, Ghada, Mona, Rahinah, Rani, Roqayyah, Walaa, Reem, Hossam, Father Hester, and Lina for your friendship.

I thank Flora, Nour, and all the daycare team for taking care of Omar. Knowing that he is in his good hands has helped me focus on my work.

I thank my parents, my brother, my in-laws and my extended family for supporting my journey through the PhD and their continued support.

I thank Ahmed and Omar for their delightful presence in my life. God bless.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Examination Committee Approval</td>
<td>2</td>
</tr>
<tr>
<td>Copyright</td>
<td>3</td>
</tr>
<tr>
<td>Abstract</td>
<td>4</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>6</td>
</tr>
<tr>
<td>List of abbreviations</td>
<td>11</td>
</tr>
<tr>
<td>List of Figures</td>
<td>12</td>
</tr>
<tr>
<td>List of Tables</td>
<td>14</td>
</tr>
<tr>
<td><strong>1 Introduction</strong></td>
<td>16</td>
</tr>
<tr>
<td>1.1 An Overview of Cognitive Radios</td>
<td></td>
</tr>
<tr>
<td>1.1.1 Spectrum Sensing</td>
<td>18</td>
</tr>
<tr>
<td>1.1.2 Spectrum Access</td>
<td>20</td>
</tr>
<tr>
<td>1.2 Thesis Objectives and Contributions</td>
<td>21</td>
</tr>
<tr>
<td>1.3 Conclusion</td>
<td>25</td>
</tr>
<tr>
<td><strong>2 Wideband Spectrum Sensing Order for Cognitive Radios</strong></td>
<td>27</td>
</tr>
<tr>
<td>2.1 Background and Literature Review</td>
<td>27</td>
</tr>
<tr>
<td>2.2 System Model</td>
<td>29</td>
</tr>
<tr>
<td>2.3 Channel Probing</td>
<td>30</td>
</tr>
<tr>
<td>2.4 Optimal Access Policy</td>
<td>32</td>
</tr>
<tr>
<td>2.5 Illustrative Numerical Results</td>
<td>36</td>
</tr>
<tr>
<td>2.6 Conclusion</td>
<td>37</td>
</tr>
<tr>
<td>2.7 Appendix</td>
<td>38</td>
</tr>
<tr>
<td><strong>3 Equal Gain Combining for Cooperative Spectrum Sensing</strong></td>
<td>39</td>
</tr>
<tr>
<td>3.1 Background and Literature Review</td>
<td>39</td>
</tr>
</tbody>
</table>
3.2 Collaborative Sensing With Equal Gain Combining
  3.2.1 Statistics of the Received Signal at the FC
  3.2.2 The MGF Approach
3.3 SEGC Under Soft Sensing and Estimation Errors
  3.3.1 Use of Soft Sensing Information
  3.3.2 Errors in Phase Estimation
3.4 Performance Analysis of SEGC
  3.4.1 Optimizing the Local and Global Thresholds
  3.4.2 A TDMA System and Throughput Comparison
  3.4.3 Comparing MRC and SEGC
  3.4.4 Impact of Phase and Synchronization Errors
3.5 Conclusion
3.6 Appendix
  3.6.1 Evaluation of MGFs under Errors in Phase Estimation
  3.6.2 The PDFs under the TDMA Scheme
  3.6.3 MRC versus EGC of Sensors’ Decisions
4 A Multiple Access Scheme for Cognitive Radios
  4.1 Background and Literature Review
  4.2 System Model and Physical Layer
    4.2.1 Primary and Secondary Power Control
    4.2.2 Impact of Imperfect Sensing
  4.3 The MAC layer
    4.3.1 SU Queueing Analysis
    4.3.2 SU Power Consumption
    4.3.3 Optimal Scheduling with the PU queue
  4.4 Queueing Delay Minimization
    4.4.1 Delay Minimization without Primary Queue
    4.4.2 Delay Minimization with Primary Queue
  4.5 Numerical Results
  4.6 Conclusions
  4.7 Appendix
5 Cognitive Throughput Through Relaying and ARQ
  5.1 Background and Literature Review
  5.2 System Model
    5.2.1 Physical Layer
# LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACK</td>
<td>ACKnowledgment</td>
</tr>
<tr>
<td>AF</td>
<td>Amplify-and-Forward</td>
</tr>
<tr>
<td>ARQ</td>
<td>Automatic Repeat ReQuest</td>
</tr>
<tr>
<td>AWGN</td>
<td>Additive White Gaussian Noise</td>
</tr>
<tr>
<td>BASTA</td>
<td>Bernoulli Arrivals See Time Averages</td>
</tr>
<tr>
<td>BS</td>
<td>Base Station</td>
</tr>
<tr>
<td>CDF</td>
<td>Cumulative Distribution Function</td>
</tr>
<tr>
<td>CLT</td>
<td>Central Limit Theorem</td>
</tr>
<tr>
<td>CR</td>
<td>Cognitive Radio</td>
</tr>
<tr>
<td>CRC</td>
<td>Cyclic Redundancy Code</td>
</tr>
<tr>
<td>DCC</td>
<td>Deflection Coefficient based Combining</td>
</tr>
<tr>
<td>DF</td>
<td>Decode-and-Forward</td>
</tr>
<tr>
<td>DFT</td>
<td>Discrete Fourier Transform</td>
</tr>
<tr>
<td>DS</td>
<td>Delay Sensitive</td>
</tr>
<tr>
<td>DT</td>
<td>Delay Tolerant</td>
</tr>
<tr>
<td>EGC</td>
<td>Equal Gain Combining</td>
</tr>
<tr>
<td>FC</td>
<td>Fusion Center</td>
</tr>
<tr>
<td>FCC</td>
<td>Federal Communications Commission</td>
</tr>
<tr>
<td>FDD</td>
<td>Frequency Division Duplexing</td>
</tr>
<tr>
<td>FDMA</td>
<td>Frequency Division Multiple Access</td>
</tr>
<tr>
<td>FFT</td>
<td>Fast Fourier Transform</td>
</tr>
<tr>
<td>KKT</td>
<td>Karush-Kuhn-Tucker</td>
</tr>
<tr>
<td>LLR</td>
<td>Log-Likelihood Ratio</td>
</tr>
<tr>
<td>MAC</td>
<td>Medium Access Control</td>
</tr>
<tr>
<td>MDC</td>
<td>Modified Deflection Coefficient</td>
</tr>
<tr>
<td>MGF</td>
<td>Moment Generating Function</td>
</tr>
<tr>
<td>MPR</td>
<td>Multiple Packet Reception</td>
</tr>
<tr>
<td>MRC</td>
<td>Maximum Ratio Combining</td>
</tr>
<tr>
<td>NACK</td>
<td>Negative ACKnowledgment</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>------------------------------</td>
</tr>
<tr>
<td>PDF</td>
<td>Probability Density Function</td>
</tr>
<tr>
<td>PR</td>
<td>Primary Receiver</td>
</tr>
<tr>
<td>PT</td>
<td>Primary Transmitter</td>
</tr>
<tr>
<td>PU</td>
<td>Primary User</td>
</tr>
<tr>
<td>QoS</td>
<td>Quality-of-Service</td>
</tr>
<tr>
<td>ROC</td>
<td>Receiver Operation Characteristics</td>
</tr>
<tr>
<td>SEGC</td>
<td>Sensing with Equal Gain Combining</td>
</tr>
<tr>
<td>SIC</td>
<td>Successive Interference Cancellation</td>
</tr>
<tr>
<td>SINR</td>
<td>Signal-to-Interference-and-Noise Ratio</td>
</tr>
<tr>
<td>SNR</td>
<td>Signal-to-Noise Ratio</td>
</tr>
<tr>
<td>SR</td>
<td>Secondary Receiver</td>
</tr>
<tr>
<td>ST</td>
<td>Secondary Transmitter</td>
</tr>
<tr>
<td>SU</td>
<td>Secondary User</td>
</tr>
<tr>
<td>TDD</td>
<td>Time Division Duplexing</td>
</tr>
<tr>
<td>TDMA</td>
<td>Time Division Multiple Access</td>
</tr>
<tr>
<td>WSN</td>
<td>Wireless Sensor Network</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

1.1  2011 US frequency allocations (3-30 GHz range) from [1] .......................... 17  
1.2  Spectrum utilization measurement example ............................................. 17  
2.1  Maximum secondary utility versus maximum collision probability .......... 36  
2.2  Variation of the average SU utility with the average probing SNR .......... 37  
3.1  Complementary receiver operation characteristics (C-ROC) curves .... 54  
3.2  C-ROC curves for N = 30 sensors for the SEGС and TDMA schemes 57  
3.3  Throughput of the SEGС and the TDMA schemes .................................... 59  
3.4  Comparison 1 between SEGС and MRC .................................................. 60  
3.5  Comparison 2 between SEGС and MRC .................................................. 61  
3.6  SEGС performance under phase errors .................................................... 62  
3.7  SEGС performance under phase and synchronization errors ................. 62  
3.8  Variation of the minimum $Q_{FA}$ with the SNR ...................................... 63  
3.9  Variation of the minimum $Q_{MD}$ with the reporting SNR ....................... 70  
4.1  Cognitive multiple access system ............................................................. 75  
4.2  Cognitive multiple access system transmission cases ........................... 82  
4.3  The birth-death Markov chain describing the SU queue ......................... 84  
4.4  Performance of the DS and DT schemes versus secondary arrival rate .... 99  
4.5  Overall packet loss rate for DT and DS schemes .................................... 100  
4.6  Variation of delay with sensing error probabilities ............................... 100  
4.7  Variation of PU outage probability with the sensing error probabilities 101  
4.8  Impact of buffer size $N$ on queueing delay in the DT scheme ............... 101  
4.9  Values of the optimal scheduling parameters ......................................... 102  
4.10 Queueing delay as a function of the SU arrival rate $\theta$ ......................... 102  
4.11 Discrepancy between simulation results and those obtained analytically 103  
5.1  Cognitive multiple access scheme employing feedback .......................... 112  
5.2  Markov chain representing the primary queue given the ARO ............... 120  
5.3  Markov chain representing initial states of PU queue for explicit ACK 120
5.4 Markov chain representing initial states of PU queue for explicit NACK

5.5 SU throughput using exact and approximate methods, explicit NACK

5.6 SU throughput using exact and approximate methods, explicit ACK

5.7 SU throughput for various values for the minimum required PU rate

5.8 Optimal access probabilities for $R^*_p = 0.94$

5.9 Optimal access probabilities for $\Delta P = 0.2$ and $R^*_p = 0.94$

5.10 SU throughput under explicit ACK and NACK schemes

5.11 Contrasting proposed MAC scheme with comparison systems

5.12 Contrasting proposed MAC scheme with a sensing-only system

5.13 SU throughput for the constrained and unconstrained delay problems

5.14 SU throughput for various values of $\tau$

5.15 SU throughput for $\theta = 0.05$

6.1 Three-message superposition coding scheme

6.2 Solution of throughput problem with partial power budget

6.3 Phase durations for throughput problem with partial power budget

6.4 Comparing problem 2 with two-hop routing

6.5 Comparison 1 between problem 2 and interference model

6.6 Comparison 2 between problem 2 and interference model

7.1 Comparing stable matching with optimal assignment

7.2 Matching under large and small interference

7.3 Users’ percentage under stable and optimal matching
LIST OF TABLES

5.1 Steady state probabilities for the PU queue Markov chain . . . . . . . . 122
7.1 A sample $3 \times 3$ PU and SU utility matrix . . . . . . . . . . . . . . . . . 184
7.2 Sample preference matrices . . . . . . . . . . . . . . . . . . . . . . . . . 185
Chapter 1

Introduction

In this introductory chapter, we will present an overview of cognitive radios and discuss their two main tasks. We will also outline the thesis objectives and contributions.

1.1 An Overview of Cognitive Radios

A cognitive radio (CR) is a radio that can change its transmit parameters in real-time based on interactions with the environment in which it operates. These parameters include for instance transmit power, carrier frequency, and modulation strategy [2]. The cognitive radio term was first coined by Mitola in 1998 and formally in [2]. Since its inception, CRs quickly gained grounds as a possible solution to the problem of spectrum scarcity along with the notion of dynamic spectrum allocation.

While the wireless spectrum is seemingly crowded, with no more possible assignments for new users or services, studies by regulatory bodies, such as the Federal Communications Commission (FCC), have shown the real problem of spectrum scarcity to be related to the fixed assignment of the radio resource [3]. Fig. 1.1 shows the 2011 US frequency allocations for the 3 – 30 GHz. While, it seems that no more assignments are possible, the authors in [4] have demonstrated a typical utilization of 0.5% in the 3-4 GHz frequency band. This utilization drops to 0.3% in the 4-5 GHz band as is clear in Fig. 1.2. The study was carried out in the downtown Berkeley
area and confirms previous studies which report temporal and geographic variations in the usage of allocated spectrum [3].

Therein comes the role of CRs with an agility that permits secondary, or unlicensed, users (SUs) to coexist with primary, or licensed, users under the constraint that the primary users (PUs) are guaranteed some performance requirements such as a maximum tolerable outage probability or a minimum acceptable throughput [5], [6]. The goal is to make the most efficient use of the wireless spectrum while preserving the hierarchical structure of the network with the PUs being the rightful owners of the spectrum and the SUs being opportunistic users that seek possible transmission opportunities.

To achieve the aforementioned goal, CRs broadly require two main tasks. First, they must be able to scan the frequency band of interest to assess the presence of active PU(s) through a spectrum sensing process. Whether they access the spectrum band concurrently with the PUs under interference constraints or only use the available spectrum holes that are left by the PUs, the SUs need spectrum sensing to quickly
and reliably discern primary activity. Second, given the sensing results, the CRs need to implement an adequate protocol for using the spectrum which we will collectively refer to as the spectrum access technique. We discuss next some details on each of these two key tasks for CRs.

### 1.1.1 Spectrum Sensing

Spectrum sensing is one of the key enabling technologies for cognitive radios. Before the CRs can begin to use the spectrum, a mechanism must be put in place to allow them to know when the primary users are active or inactive. There are a variety of ways through which spectrum sensing can be carried out. Energy detection, matched filtering and cyclostationary feature detection can be considered as the most commonly used ways to detect primary signals [7].

Among the above methods, however, energy detection is considered the simplest and most practical to implement. An energy detector works by measuring the energy received on a primary band during an observation interval and declares a white space if the measured energy is less than a properly set threshold [7]. While the main disadvantage of this method is that it cannot discriminate between sources of received energy, such as the primary signal, noise or other cognitive users, especially in low signal-to-noise ratio (SNR) [8], it remains a popular mechanism for spectrum sensing in the CR literature. Because of its low cost and simplicity of implementation, energy detection is also the employed spectrum sensing technique in this thesis.

In reality, the spectrum sensing process is usually cast as a binary hypothesis test, where \( \mathcal{H}_0 \) denotes an idle PU, while \( \mathcal{H}_1 \) denotes an active one. Using an energy detector, the test statistic is [9]

\[
T = \sum_{n=1}^{N} |Y(n)|^2,
\]  

where \( Y(n) \) is the received signal.
where \( N \) is the number of samples taken by the CR sensor, this can also be considered as a proxy for the overall sensing duration, while \( Y(n) \) is the \( n \)th sample of the received signal at the sensor and can be expressed as,

\[
Y(n) = \begin{cases} 
W(n) & \text{under } \mathcal{H}_0 \\
h_{sc} X(n) + W(n) & \text{under } \mathcal{H}_1,
\end{cases}
\] (1.2)

where \( W(n) \) is the receiver noise, while \( X(n) \) is the primary signal, under \( \mathcal{H}_1 \), and \( h_{sc} \) is the sensing channel gain between the PU and the SU. If the sensing threshold is \( \kappa \), then the resulting probabilities of false alarm and miss-detection can be defined, respectively, as

\[
p_{FA} = \Pr \{ T > \kappa \mid \mathcal{H}_0 \},
\] (1.3)

\[
p_{MD} = \Pr \{ T < \kappa \mid \mathcal{H}_1 \},
\] (1.4)

It is noted that the sensing time, or \( N \), and the sensing threshold \( \kappa \) determine the resulting detection probabilities. One may choose to increase the sensing duration so as to achieve a high confidence in the resulting statistic, however, this would come at the expense of less time available for CR transmission resulting in a sensing-throughput tradeoff that is characteristic of CR networks and which has been reported in [10]. Also, setting the sensing threshold too low, may mean more protection for the PU, but less transmission opportunities for the SU as the false alarm probability will become quite high. Hence, the sensing parameters, whether the sensing threshold or duration, need to be carefully chosen to benefit the CR without degrading the performance of the Primary network. A design objective which we undertake in this thesis.

Furthermore, many communication scenarios exist where a direct sensing chan-
nel may not exist between the PU and the SU due to physical obstructions in the surrounding environment. Spectrum sensing in such case becomes impossible and well-known difficulties such as the hidden-terminal problem arise whereby a CR may not be aware of the activity of a PU that is within close vicinity, but cannot be detected because of physical barriers. The SU may then declare the PU to be idle and start transmission, thereby causing interference to the primary network. To achieve a specified detection performance, despite heavy shadowing and fading, cooperative spectrum sensing was proposed as an effective way to reliably detect primary activity [11–13]. Cooperation is achieved by allowing different SUs to share their sensing results, usually via a central node, which makes a global decision on the occupancy status of the licensed band. Traditionally, cooperative sensing is carried out in a way through which the CRs orthogonally communicate their sensing result to the central node. This leads to significant sensing delays or bandwidth loss. Cooperative sensing should be performed efficiently and reliably in order to achieve the goals of the cognitive network. This is yet another objective that we undertake and which is detailed in the next section. Before that, however, we discuss the specifics of the spectrum access techniques investigated in this work.

1.1.2 Spectrum Access

How the SU accesses the primary channel on the basis of its observations is an active area of research with major open issues and challenges. To date, it is possible to roughly group spectrum access techniques into three broad categories [14]:

- Under the *interweave* paradigm, CRs exploit knowledge about the activity of the PU to access the channel *only* when the PU is deemed inactive.

- Under the *underlay* paradigm, it is assumed that the CR is aware of the interference it causes to the PU, so that concurrent transmission with the PU may
occur only if the interference caused to the PU is kept at an agreed-upon level or a so called interference temperature level [15].

- Under the overlay paradigm, the PU and the SU are allowed to concurrently transmit within the same frequency band provided that certain quality of service parameters are guaranteed to the PU such as a minimum rate or a maximum outage probability [16–18].

It is noted above that the overlay approach does require knowledge on the part of the PU about the presence of the SU. While some works in the open literature advocate strict transparency of the secondary network to primary operation, such transparency cannot be maintained given the realities of spectrum sensing and given the potential benefit that the primary network can reap from the key concept of cooperation. In fact, a primary network playing an active role in spectrum management has been shown in previous work in the literature to benefit both users [19][20]. Hence, we focus in this thesis on spectrum access schemes whereby the PUs are not oblivious to the presence of the SUs and play an active and cooperative role to maximize their benefit from the spectrum access. We consider two scheduling problems, a superposition coding cognitive relaying scheme and a spectrum access technique that involves multiple primary and secondary users. The details of each of these schemes are provided next.

1.2 Thesis Objectives and Contributions

The broad objective of this thesis is to provide novel enhanced spectrum sensing and access techniques that are shown to outperform existing schemes in the literature. Below we provide more details on the objectives of each chapter and its contribution.

In chapter 2 an SU seeks to transmit by sequentially sensing statistically independent PU channels. If a channel is sensed free, it is probed to estimate the SNR
between the SU transmitter-receiver pair over the channel. We jointly optimize the channel sensing time, the sensing decision threshold, the channel probing time, together with the channel sensing order under imperfect synchronization between the PU and the SU. The sensing and probing times and the decision threshold are assumed to be the same for all channels. We maximize a utility function related to the SU throughput under the constraint that the collision probability with the PU is kept below a certain value and taking sensing errors into account. We illustrate the optimal policy and the variation of SU throughput with various system parameters.

In chapter 3, a novel cooperative spectrum sensing technique for cognitive radio networks, termed sensing with equal gain combining (SEGC), is proposed. Cognitive radios simultaneously transmit their sensing results to the fusion center (FC) over multipath fading reporting channels. The cognitive radios estimate the phases of the reporting channels and use those estimates for coherent combining of the sensing results at the FC. A global decision is made at the FC by comparing the received signal with a threshold. We obtain the global detection probabilities and secondary throughput exactly through a moment generating function approach. We verify our solution via system simulation and demonstrate that the Chernoff bound and central limit theory approximation are not tight. The cases of hard sensing and soft sensing are considered and we provide examples in which hard sensing is advantageous to soft sensing. We contrast the performance of SEGC with maximum ratio combining of the sensors’ results and provide examples where the former is superior. Furthermore, we evaluate the performance of SEGC against existing orthogonal reporting techniques such as time division multiple access (TDMA). SEGC performance always dominates that of TDMA in terms of secondary throughput. We also study the impact of phase and synchronization errors and demonstrate the robustness of the SEGC technique against such imperfections.

In chapter 4, we study a time-slotted multiple-access system with a PU and an
SU sharing the same channel resource. The SU senses the channel at the beginning of the slot. If found free, it transmits with probability one. If busy, it transmits with a certain access probability that is a function of its queue length and whether it has a new packet arrival. Both users, the PU and the SU, transmit with a fixed transmission rate by employing a truncated channel inversion power control scheme. We consider the case of erroneous sensing. The goal of the SU is to optimize its transmission scheduling policy so as to minimize its queueing delay under constraints on its average transmit power and the maximum tolerable primary outage probability caused by the miss-detection of the PU. We consider two schemes regarding the secondary’s reaction to transmission errors. Under the so-called delay-sensitive (DS) scheme the packet received in error is removed from the queue to minimize delay, whereas under the delay-tolerant (DT) scheme, said packet is kept in the buffer and is retransmitted till correct reception. Using the latter scheme, there is a probability of buffer loss that is also constrained to be lower than a certain specified value. We also consider the case when the PU maintains an infinite buffer to store its packets. In the latter case, we modify the SU access scheme so as to guarantee stability of the PU queue. We show that the performance changes significantly if the realistic situation of a primary queue is considered. In all cases, although the delay minimization problem is non-convex, we show that the access policies can be obtained efficiently using linear programming and grid search over one or two parameters.

In chapter 5, we consider a spectrum sharing system, where the primary terminal operates in a time slotted fashion and is active only when it has a packet to send. The secondary terminal uses spectrum sensing results and the primary automatic repeat request (ARQ) feedback to access the channel probabilistically. To enhance the primary’s system performance, the secondary user acts as a relay for the primary user in the event of transmission failure on the direct link of the latter. Closed-form expressions for the primary and secondary throughputs are obtained for the described
scheme. The optimal medium access probabilities are then obtained via maximizing the secondary throughput subject to constraints that guarantee the stability of the considered queues, a minimum primary throughput and a maximum primary queueing delay. The results clearly indicate the benefits of cognitive relaying in enhancing the throughput performance for both the primary and the secondary users. Furthermore, by guaranteeing minimum rate and maximum delay requirements, our scheme is shown to provide a definitive notion of protection for the licensed users of the network.

In chapter 6 we propose a three-message superposition coding scheme in a cognitive radio relay network exploiting active cooperation between primary and secondary networks. The PU is motivated to cooperate by substantial benefits it can reap from this access scenario. Specifically, the time resource is split into three transmission phases: The first two phases are dedicated to primary communication, while the third phase is for the secondary’s transmission. We formulate two throughput maximization problems for the secondary network subject to primary user rate constraints and per-node power constraints with respect to the time durations of primary transmission and the transmit power of the primary and the secondary users. The first throughput maximization problem assumes a partial power constraint such that the secondary power dedicated to primary cooperation, i.e. for the first two communication phases, is fixed a priori. In the second throughput maximization problem, a total power constraint is assumed over all three phases of communication. We derive analytical solutions to both problems under various conditions on the direct and relaying channels and validate our findings by numerical optimization. Given the partial and total power constraints, the analytical solutions of the throughput problems can be drastically different. For both problems, we demonstrate significant throughput gains for both the primary and the secondary users through this active cooperation scheme. We find that most of the throughput gains come from minimizing the second phase
transmission time because of secondary cooperation. Finally, we demonstrate the superiority of our proposed scheme compared to a number of reference schemes that include best relay selection, two-hop routing and an interference channel model.

In chapter 7, we consider a network comprised of multiple PUs and multiple SUs. The goal of the SUs is to gain access to orthogonal channels each occupied by one PU. Only one SU is allowed to coexist with a given PU. We propose a three-stage distributed matching algorithm to pair the network users where a Stackelberg game model is assumed for the interaction between the paired PU and SU. The selected SU is given access in exchange for monetary compensation to the PU. The PU also increases its power to meet its rate requirement despite the SU’s presence. The PU seeks to optimize the interference price it offers to a given SU and the power allocation to maintain communication. The SU, on the other hand, seeks to optimize its power demand so as to maximize its utility. We show that our algorithm provides a unique stable matching. Numerical results indicate the superiority of the proposed algorithm to a random pairing of the users. Furthermore, the algorithm is shown to perform close to an optimal central controller, albeit at a reduced computational complexity.

Finally, in chapter 8, we summarize the work that we have done and illustrate possible extensions.

1.3 Conclusion

Cognitive radio technology and dynamic spectrum allocation have the potential to resolve spectrum scarcity and to maximize spectral efficiency [21]. A cognitive radio works by first scanning the wireless medium for spectrum opportunities through a spectrum sensing process, the simplest of which is energy detection. Once a transmission opportunity arises, then detailed communication techniques must be set up to define how the cognitive radio is to access the spectrum. In this chapter, we provided
an overview of spectrum sensing and spectrum access schemes that we have devised to enhance these two processes in cognitive radio systems.
Chapter 2

Wideband Spectrum Sensing
Order for Cognitive Radios with Sensing Errors and Channel SNR Probing Uncertainty

2.1 Background and Literature Review

We consider a scenario where an SU sequentially searches a number of wideband primary channels in order to identify a transmission opportunity. Once a channel is sensed free, the SU attempts to estimate its SNR and potentially uses this sensed-free channel for transmission. The cognitive user needs to balance the requirements of operational reliability and throughput maximization. Indeed, if the time dedicated to sensing and SNR probing is increased, the cognitive terminal gets a more reliable sensing result and a better SNR estimate. However, this reduces the time available for actual data transmission. Thus, a key question arises regarding the optimal sequence in which the channels should be sensed, the optimal sensing parameters, and the

---

1Each of the primary channels is assumed to be wideband. This places a hardware constraint on the SU and necessitates sequential search.
optimal SNR probing parameters.

It was shown in [22] that the optimal sensing sequence in the homogeneous channel capacity case is to sort the channels in ascending order of the ratio between the channel sensing time, which is considered to be fixed, and the channel availability probability. In [23], the authors investigate the optimal channel selection problem assuming error-free sensing. The transmitter’s objective is to choose the strategy that maximizes transmission reward minus the probing costs. The authors of [24] extend the work in [23] by considering sensing errors. Although the optimization of the sensing time is included in [24], the channel sensing order is random and channel probing is assumed perfect. Furthermore, the impact of sensing errors on the primary network is not investigated.

Besides, [25] and [26] employ adaptive-rate transmission and obtain the sensing order via optimizing a secondary utility function. Spectrum sensing and the optimization of its parameters are incorporated rigorously in [26]. However, two key issues are neglected, namely, channel SNR probing is assumed to be perfect and instantaneous and the collision probability with the PU is unconstrained.

Tackling the problem of determining the optimal sensing sequence and parameters and the optimal probing parameters, while taking into account the above-mentioned limitations and imperfect synchronization between the PU and the SU, the contributions of this chapter are threefold: i) incorporating channel SNR probing into the secondary utility function; ii) jointly optimizing the channel sensing sequence, spectrum sensing parameters and channel probing duration, and iii) constraining the maximum allowable collision probability with the primary network due to miss-detection.
2.2 System Model

The secondary transmitter attempts to access one of $K$ channels of a primary network. The PUs’ activity follows a time-slotted structure to which the SU is synchronized. Synchronization is imperfect and we account for this using two parameters $\vartheta_s$ and $\vartheta_b$, where $\vartheta_s$ denotes a start time that the SU waits before sensing and probing the channels, while $\vartheta_b$ denotes a backoff time that the SU uses to vacate the channel or, if not yet transmitting, to cease trying to gain access to the PU channels for fear of entering into the next time slot. We assume $\vartheta_s$ and $\vartheta_b$ are known to the SU on the basis of the synchronization algorithm that it uses. During a slot duration, $T$, a particular channel is either used by the PU or is vacant. The probability of channel $i$ being free from primary activity is denoted as $\nu_i$. The values of $\nu_i$ ($1 \leq i \leq K$) are assumed to be known a priori to the SU.\footnote{These probabilities may be reliably estimated before the algorithm is implemented.} The primary activity over a particular channel is independent from one slot to another and is also independent from other channels’ activities. The channel gains are assumed to be fixed over the time slot and change independently from a given slot to another. Further, the channel gains across the bands are assumed independent. The channels are wideband and are sensed by the SU one channel at a time. A question arises as to the best channel sensing sequence so as to optimize some performance metric.

The SU also optimizes the sensing and probing durations. The sensing duration, $\tau_s$, is the time used to sense primary activity, whereas the probing duration, $\tau_p$, is the time used by the SU to probe the SNR of a sensed-free channel. Increasing $\tau_s$ enhances the PU detection reliability. However, it also decreases the time available for data transmission. Similarly, increasing $\tau_p$ increases the accuracy of the channel’s SNR estimate. This comes at the price of reducing time available for SU transmission.

For sensing, the SU employs an energy detector to sense the channel state. The detector averages the received energy over a number of consecutive samples, $N$, and
compares the average to a detection threshold, $E_T$. If the energy detector’s sampling frequency is $f_s$, the time taken to sense a channel is $\tau_s = N/f_s$. A false alarm occurs if a free channel is sensed to be busy, thereby causing the SU to lose a transmission opportunity. A miss-detection, on the other hand, makes a busy channel sensed free. If the SU transmits, the PU and SU transmissions will collide. Miss-detection events harm the primary network and degrade its throughput. The false alarm and miss-detection probabilities are functions of $E_T$ and $N$, as well as the parameters of the sensing channel between the primary transmitter and the secondary transmitter.

2.3 Channel Probing

Once a channel is sensed free, SNR probing proceeds by sending $M$ pilots from the SU receiver to the SU transmitter, which computes an SNR estimate, $\hat{\Omega}$, and uses it to determine its transmission rate. If $f_{s,p}$ is the probing sampling frequency, then $\tau_p = M/f_{s,p}$. Channel probing is impacted by thermal noise and by the interference from the primary transmitter if it is ON.

Herein, we consider a typical I/Q receiver and $M$ pilots. The $M$ measurements are averaged so the received in-phase and quadrature components are given by:

$$A_I = \sqrt{\Omega \sigma_S^2 \cos \phi + \varpi \sqrt{P_{PS}} \cos \phi'} + n_I,$$

$$A_Q = \sqrt{\Omega \sigma_S^2 \sin \phi + \varpi \sqrt{P_{PS}} \sin \phi'} + n_Q,$$

(2.1)

where $\Omega$ is the SNR of the channel between the secondary transmitter-receiver with no interference from the PU, $\sigma_S^2$ is the noise variance, so $\Omega \sigma_S^2$ is the power of the pilot symbol, $\phi$ is the phase difference between the received and locally generated carriers. Parameter $\varpi = 0$ if the primary is inactive and $\varpi = 1$ if it is ON. $P_{PS}$ is the power from the primary transmitter to the secondary receiver, and $\phi'$ is the phase difference.

---

3Channel reciprocity is assumed, i.e. the channel is the same in either direction of communication between the SU transmitter-receiver.
between the received PU interference signal and the locally generated carriers. The noise terms $n_I = \frac{1}{M} \sum_{m=1}^{M} n_{I,m}$ and $n_Q = \frac{1}{M} \sum_{m=1}^{M} n_{Q,m}$, with $n_{I,m}$ and $n_{Q,m}$ being the thermal noise components added to the $m$th sample with each component being a real zero-mean Gaussian random variable with variance $\sigma_S^2/2$. Both $n_I$ and $n_Q$ are thus zero-mean Gaussian random variables with variance $\sigma_S^2/2M$. With the exception of the thermal noise, all other parameters are assumed to be constant during the time slot and, hence, during the probing period which constitutes a fraction of the slot duration. The probability that $\omega = 0$ when the channel is sensed to be free can be readily obtained using Bayes theorem:

$$\Pr\{\omega = 0 \mid \text{channel free}\} = \frac{(1 - p_{\text{FA}}) \upsilon_i}{(1 - p_{\text{FA}}) \upsilon_i + p_{\text{MD}} (1 - \upsilon_i)}, \quad (2.2)$$

where $p_{\text{FA}}$ and $p_{\text{MD}}$ are the false alarm and miss-detection probabilities, respectively. To simplify notation, we will write $\Pr\{\omega = 0 \mid \text{channel free}\}$ as $\Pr\{\omega = 0\}$.

The SNR estimate, $\hat{\Omega}$, is a function of both $A_I$ and $A_Q$. It can be obtained via the maximum likelihood (ML) principle. Specifically, $\hat{\Omega} = \arg\max_{\Omega} f(A_I, A_Q \mid \Omega)$. In Appendix 2.7, we provide expressions for $f(A_I, A_Q \mid \Omega)$ and the other relevant probability distribution functions.

The SU uses adaptive modulation, i.e. the transmission rate is adjusted according to the channel’s SNR. The secondary transmission rate, given in bits/channel use is $\log(1 + g(\hat{\Omega}))$, where $g(\hat{\Omega})$ is a function of the observed SNR $\hat{\Omega}$. This function is chosen to maximize the expected throughput. If the actual channel SNR is $\Omega$, the packet is delivered successfully to the secondary receiver only when $\Omega \geq g(\hat{\Omega})$ if the PU is OFF and when $\Omega \geq g(\hat{\Omega}) \left(1 + \frac{P_{\text{PS}}}{\sigma_S^2}\right)$ if the PU is active. This is because the condition of correct reception during concurrent primary/secondary operation is $\frac{\Omega \sigma_S^2}{P_{\text{PS}} + \sigma_S^2} \geq g(\hat{\Omega})$, i.e. that the actual channel capacity is greater than or equal to the used rate.

We choose $g(\hat{\Omega})$ to maximize the product of the rate $\log \left(1 + g(\hat{\Omega})\right)$ with the probability

\[\text{Recall that the prior probability that the channel i is free is } \upsilon_i.\]

\[\text{Recall that } \Omega \text{ is defined as the SNR of the channel between the secondary transmitter-receiver with no interference from the PU.}\]
of the packet being received successfully, i.e. the function:

\[
\log \left(1 + g(\hat{\Omega}) \right) \left[ \Pr \{\varpi = 0 | \hat{\Omega} \} \int_{g(\hat{\Omega})}^{\infty} f \left( \Omega | \hat{\Omega}, \varpi = 0 \right) d\Omega \\
+ \Pr \{\varpi = 1 | \hat{\Omega} \} \int_{g(\hat{\Omega})}^{\infty} f \left( \Omega | \hat{\Omega}, \varpi = 1 \right) d\Omega \right],
\]

(2.3)

where \( f(\Omega|\hat{\Omega}, \varpi) \), given in the Appendix, is the posterior probability of the channel’s SNR given \( \varpi \) and the estimate \( \hat{\Omega} \). If \( g(\hat{\Omega}) \) is increased, this means an increase in the SU transmission rate. However, the probability of correct packet reception decreases as there is a higher probability the true channel SNR is lower than \( g(\hat{\Omega}) \) causing the channel capacity to drop below the transmission rate. We obtain \( g(\hat{\Omega}) \) by numerically searching for its value to maximize (2.3) given \( \hat{\Omega} \).

### 2.4 Optimal Access Policy

Each time slot duration is comprised of stages equal to the number of available channels. Given the list of channels to sense, \( s_1, s_2, \ldots, s_K \), the cognitive user starts sensing \( s_1 \). If the channel is found busy, the SU proceeds to \( s_2 \). If the channel is sensed free, the SU probes the channel to estimate the SNR. It computes the expected reward if it transmits and compares it with the expected reward if the channel is skipped. If the reward for transmission exceeds the reward for skipping, the channel is used for transmission, otherwise it is skipped and \( s_3 \) is sensed. A skipped channel cannot be recalled at a later stage. The expected reward of using a free channel which is effectively sensed to be free, and after \( k \) channel sensing and \( m \) channel probing durations is

\[
R(k, m, \hat{\Omega}) = \left(1 - \frac{k\tau + m\tau_p + \vartheta_s + \vartheta_b}{T}\right)^+ \times \log \left(1 + g(\hat{\Omega}) \right) \int_{g(\hat{\Omega})}^{\infty} f \left( \Omega | \hat{\Omega}, \varpi = 0 \right) d\Omega,
\]

(2.4)
where \((x)^+ = \max\{x, 0\}\). The first term accounts for the time slot fraction used for spectrum sensing and channel probing. Parameter \(m \leq k\) because a channel is probed only if it is sensed to be free. If the channel is sensed to be free while it is busy, the expected reward is

\[
\bar{R}(k, m, \hat{\Omega}) = \left(1 - \frac{k\tau + m\tau_p + \vartheta_s + \vartheta_b}{T}\right)^+ \times \log \left(1 + g(\hat{\Omega})\right) \int_0^\infty \frac{g(\theta)}{(1 + \frac{\vartheta_p}{\vartheta_s})} \left(\Omega|\hat{\Omega}, \varpi = 1\right) d\Omega.
\]  

(2.5)

The expected rate is therefore given by

\[
R^*(k, m, \hat{\Omega}) = \Pr\{\varpi = 0|\hat{\Omega}\} R(k, m, \hat{\Omega}) + \Pr\{\varpi = 1|\hat{\Omega}\} \bar{R}(k, m, \hat{\Omega}).
\]  

(2.6)

The secondary utility function \(U(k, m)\), is the expected reward at the \(k\)th sensing stage if \(m\) channels have been previously probed, \(m \in \{0, 1, \ldots, k - 1\}\). This utility contains three parts: \(U(k, m) = Z_1 + Z_2 + Z_3\) which we explain below.

First, the utility when the channel is free and is sensed to be free, which happens with probability \(v_{sk} (1 - P_{FA})\), is

\[
Z_1 = v_{sk} (1 - P_{FA}) \times \left[ \mathbb{E}\left\{ R\left(k, m + 1, \hat{\Omega}\right) 1\left(R^*\left(k, m + 1, \hat{\Omega}\right) \geq U(k + 1, m + 1)\right) \right\} + U(k + 1, m + 1) \Pr\{R^*\left(k, m + 1, \hat{\Omega}\right) < U(k + 1, m + 1)\}\right].
\]  

(2.7)

where expectation \(\mathbb{E}\) operates over the measured SNR, \(\hat{\Omega}\). The first term in the brackets gives the utility when the channel is used for transmission, whereas the second term is the utility if the channel is skipped because the expected reward from its immediate use is lower than the utility accrued by skipping it and sensing the next channel. Further, \(1\) is the indicator function. The utility when the channel is sensed to be busy and, hence, skipped, is given by

\[
Z_2 = \left[ v_{sk} P_{FA} + (1 - v_{sk}) (1 - P_{MD}) \right] U(k + 1, m).
\]  

(2.8)

If the channel is sensed to be free while it is busy, then the cognitive terminal probes the channel and decides whether to transmit or skip depending on the SNR. The
corresponding utility is
\[ Z_3 = (1 - v_{sk}) p_{MD} \times \left[ \mathbb{E}\{ R(k, m + 1, \hat{\Omega}) \mathbb{I}\{ R^* (k, m + 1, \hat{\Omega}) \geq U(k + 1, m + 1) \} \right] \]
\[ + U(k + 1, m + 1) \Pr\{ R^* (k, m + 1, \hat{\Omega}) < U(k + 1, m + 1) \} \] . \tag{2.9}

If the SU reaches the last channel, skipping the latter yields zero reward. If this channel is sensed free, the SU, as an opportunistic user, would always transmit regardless of the SNR value. The expected reward at this stage is given by,
\[ U(K, m) = v_{sk} (1 - p_{FA}) \mathbb{E}\{ R(K, m + 1, \hat{\Omega}) \} + (1 - v_{sk}) p_{MD} \mathbb{E}\{ R(K, m + 1, \hat{\Omega}) \} . \tag{2.10} \]

The utility function can be obtained recursively starting from (2.10). While maximizing its utility, the SU should also be aware of the loss of throughput it causes to the primary network. This is reflected in the probability of collision given by\(^6\)
\[ p_c = \sum_{k=1}^{K-1} \sum_{m=0}^{k-1} \Pr\{ R^* (k, m + 1, \hat{\Omega}) \geq U(k + 1, m + 1) \} \]
\[ \times (1 - v_{sk}) p_{MD} \Gamma_{k,m} + \sum_{m=0}^{K-1} (1 - v_{sk}) p_{MD} \Gamma_{K,m}, \tag{2.11} \]

where \( \Gamma_{k,m} \) is the probability of reaching the \( k \)th stage with \( m \) probed channels. Note that \( \Gamma_{1,0} = 1 \) because the secondary terminal always starts at the first stage with no previously probed channels. Probability \( \Gamma_{k,m} \) can be computed recursively as follows for \( k \geq 2 \) and \( m \in \{1, 2, \cdots, k-2\} \): 
\[ \Gamma_{k,m} = \Gamma_{k-1,m} \left[ v_{sk-1} p_{FA} + (1 - v_{sk-1}) (1 - p_{MD}) \right] \]
\[ + \Gamma_{k-1,m-1} \left[ v_{sk-1} (1 - p_{FA}) + (1 - v_{sk-1}) p_{MD} \right] \times \Pr\{ R^* (k - 1, m, \hat{\Omega}) < U(k, m) \} . \tag{2.12} \]

Furthermore, for \( m = 0 \) we have
\[ \Gamma_{k,0} = \Gamma_{k-1,0} \left[ v_{sk-1} p_{FA} + (1 - v_{sk-1}) (1 - p_{MD}) \right] \tag{2.13} \]

\(^6\) If the SU miss-detects the PU’s presence and starts probing and transmitting on the channel, this will lead to a collision with the PU packet. Probing that is not followed by data transmission is considered to have a negligible impact on the PU. Including its effect is straightforward and is simply done by removing the \( \Pr\{\} \) term in (2.11).
and for \( m = k - 1 \),

\[
\Gamma_{k,k-1} = \Gamma_{k-1,k-2} \left[ v_{s_{k-1}} (1 - p_{FA}) + (1 - v_{s_{k-1}}) p_{MD} \right] \times \Pr \left\{ R^* \left( k - 1, k - 1, \hat{\Omega} \right) < U (k, k - 1) \right\}.
\]

(2.14)

Considering the objective of maximizing the secondary utility while constraining the collision probability with the primary network, the optimal sensing and probing parameters are obtained offline by computing \( g(\hat{\Omega}) \) and then solving the following problem,

\[
\maximize_{N, E_T, M, S} U (1, 0)
\]

subject to: \( p_c \leq p_{c,\text{max}} \),

where \( S \) is the set of all possible channel sequences and \( p_{c,\text{max}} \) is the maximum allowed collision probability.\(^7\) Note that \( N \) takes integer values in the range \([1, \lfloor T \cdot f_s \rfloor]\) while \( M \) takes the range \([1, \lfloor T \cdot f_{s,p} \rfloor]\), where \( \lfloor x \rfloor \) is the maximum integer less than \( x \). Also, \( N/f_s + M/f_{s,p} \leq T \). Using [25], the complexity for finding \( S \) in the offline algorithm can be made to be \( O(K.2^K-1) \). The problem in (2.15), which is non-convex in the remaining variables, is then solved exhaustively where a quantization over \( E_T \) is used. Hence, the complexity of the offline algorithm can be obtained by multiplying the aforementioned complexity by the range of values for \( N, M, \) and the number of values for the discretized variable \( E_T \). Once the optimal values are obtained, the online secondary operation proceeds as summarized in Algorithm 2. The offline algorithm is computationally intense, however it needs only to be computed when the statistics of the wireless channel change. On the other hand, the worst case complexity of the online algorithm is \( O(K) \).

\(^7\)Recall that \( N = \tau_s f_s \) and \( M = \tau_p f_{s,p} \).
Algorithm 1 Online implementation of the sequential sensing given optimal $N$, $E_T$, $M$, $S$ and $U(k, m)$.

1: **Initialization:** $k = 1$, $m = 0$, timer $t = 0$. 
2: **While** $k < K$: sense channel $s_k$, $t = t + \tau_s$. If $s_k$ is sensed free, probe it to obtain $\hat{\Omega}$ and set $t = t + \tau_p$, otherwise set $k = k + 1$. If channel is probed, calculate $R^*(k, m + 1, \hat{\omega})$. If $R^*(k, m + 1, \hat{\omega}) \geq U(k + 1, m + 1)$, then transmit on channel $s_k$ and break, otherwise set $m = m + 1$ and $k = k + 1$. If $t > T - (\tau_s + \tau_p + \vartheta_b)$, break. 
3: **If** $k = K$: sense $s_K$. If channel $s_K$ is sensed free, probe it to get $\hat{\Omega}$ and transmit, otherwise wait for the next time slot.

2.5 Illustrative Numerical Results

We present results here for the case of a Z-interference channel, i.e. $P_{PS} = 0$. The false alarm and miss-detection probabilities are as in [26]: $p_{FA} = 1 - \Gamma_{inc}\left(\frac{NE_T}{\sigma_S^2}, N\right)$ and $p_{MD} = \Gamma_{inc}\left(\frac{NE_T}{\sigma_S^2 + \epsilon}, N\right)$, where $\Gamma_{inc}$ is the incomplete Gamma function and $\tau$ is the sensing channel average SNR.

We assume that $\Omega$ is exponentially distributed.

![Figure 2.1](image_url) Simulation results for $K = 4$ primary channels. The $\nu_i$’s are [.9058,.127,.9134,.6324], $\bar{\epsilon} = 1.5$ dB, the average probing SNR $\overline{\Omega} = 3$ dB, $\sigma_S^2 = 1$, $f_s = f_{s,p} = \frac{1}{20}$. We consider $p_{c,max} = [.005,.02,.05,.08]$. The optimal channel sequence is $\{s_1, s_2, s_3, s_4\} = \{3,1,4,2\}$ for all the $p_{c,max}$ values except for $p_{c,max} = 0.005$ where $\{s_1, s_2, s_3, s_4\} = \{3,1,2,4\}$.

Fig. 2.1 shows the values of $U(1,0)$ with $p_{c,max}$ under perfect synchronization, i.e. $\vartheta_s = \vartheta_b = 0$. As expected, the utility of the SU increases as $p_{c,max}$ increases.

Further, we compare our scheme against a random scheduling scheme where the $\vartheta_s$ and $\vartheta_b$ are assumed to be uniformly distributed. These expressions presume that the sensing channel is a Rayleigh fading channel with a known distribution, but whose exact gain at any time slot is not known to the secondary transmitter.
channel sensing order is chosen arbitrarily. The severe degradation in the random scheduling is due to the fact that not all channel sensing sequences yield the required $p_{c, \text{max}}$ and, hence, lead to a zero utility. In Fig. 2.2 we consider a comparison of the average reward obtained from the online algorithm for various values of the synchronization errors. It is clear that the SU loses throughput as the synchronization errors increase at the expense of protecting the PU.

### 2.6 Conclusion

This chapter considered the joint optimization of channel sensing sequence, spectrum sensing parameters and channel probing duration in wideband channels under imperfect synchronization between the PU and the SU, sensing errors, and SNR probing uncertainty. An SU seeks a transmission opportunity by sequentially sensing $K$ statistically independent PU channels. If a channel is sensed free, it is probed to estimate the SNR between the SU transmitter-receiver pair over the channel. We provided optimal solutions for the problem under a collision constraint with the PU data, and investigated the variation of the secondary’s throughput with various system parameters.
2.7 Appendix: SNR Estimation

Since \( n_I \) and \( n_Q \) are normally distributed, the conditional distribution of the in-phase and quadrature components \( f(A_I, A_Q | \Omega, P_{PS}, \phi, \phi', \varpi) \) is a bivariate Gaussian distribution. To estimate \( \Omega \), we use the Bayesian approach to average out the independent nuisance parameters \( \phi, \phi', P_{PS} \), assuming their probability distributions are known so that we obtain \( f(A_I, A_Q | \Omega, \varpi) \). Given \( f(A_I, A_Q | \Omega, \varpi) \) and assuming \( f(\Omega) \) is known, we can use the law of total probability and equation (2.2) to obtain the distributions \( f(A_I, A_Q | \Omega) \), \( f(A_I, A_Q | \varpi) \) and \( f(A_I, A_Q) \).

Using transformation of random variables, we can obtain \( f(\hat{\Omega} | \Omega, \varpi) \), which can in turn be used to obtain the following:

\[
f(\Omega | \hat{\Omega}, \varpi) = \frac{f(\hat{\Omega} | \Omega, \varpi) f(\Omega)}{\int_0^\infty f(\hat{\Omega} | \Omega, \varpi) f(\Omega) d\Omega}, \tag{2.16}
\]

\[
Pr\{\varpi | \hat{\Omega}\} = \frac{Pr\{\varpi\}}{f(\hat{\Omega})} \int_0^\infty f(\hat{\Omega} | \Omega, \varpi) f(\Omega) d\Omega, \tag{2.17}
\]

\[
f(\hat{\Omega}) = \sum_{i=0,1} Pr\{\varpi = i\} \int_0^\infty f(\hat{\Omega} | \Omega, \varpi) f(\Omega) d\Omega. \tag{2.18}
\]

\(^9\)Random variables \( \phi \) and \( \phi' \) are usually assumed to be uniformly distributed over the interval \([0, 2\pi] \).
Chapter 3

Equal Gain Combining for Cooperative Spectrum Sensing in Cognitive Radio Networks

3.1 Background and Literature Review

Cooperative sensing is achieved by allowing different secondary users to share their sensing results, via a central node or fusion center (FC), which makes a global decision on the occupancy status of the licensed band.

The local sensors typically transmit sensing information to the FC through orthogonal channels, for example using time division multiple access (TDMA) or frequency division multiple access (FDMA). The TDMA approach does not suit the cognitive setting because significant delay and throughput loss occur as the number of sensors increase. On the hand, the use of FDMA through sending the decisions on orthogonal frequency bands requires a large bandwidth [27]. These issues contradict the basic premise of cognitive radios in which the terminals search for transmission opportunities and can only communicate with the FC through a low-rate reporting channel with a limited bandwidth.
In order to address this issue of efficiently reporting the sensing results to the FC, simultaneous transmission by the local sensors was considered in [28–31]. In [28], the FC collects data in multiple slots, each involving a random number of transmitting sensors. Specifically, sensors with the same data value transmit (if they decide to do so) using the same waveform on a multi-access fading channel, and the design criterion is to find the optimal mean transmission rate so that the detection error exponent is maximized. The use of orthogonal waveforms eliminates interference among sensors with different data values and makes it possible to have a coherent combining of transmissions. This, however, happens only in the absence of channel fading. In [29], each sensor makes a binary local decision and communicates it to the FC simultaneously with the other sensors. The authors investigate the detection performance in terms of the error probability and error exponent for Rayleigh and Rician fading scenarios. The communication is assumed to be non-coherent, meaning that the channel gains are unknown at both the sensors and the FC. Furthermore, detection errors, whether a false alarm or a miss-detection, are therein assumed to have an equal impact and henceforth combined in a single error term, which may not suit cognitive networks since the two users, namely the primary and the secondary, will be impacted differently by the false alarm and miss-detection events.

Reference [30] also considers the simultaneous transmission of sensors’ decisions in a wireless sensor network (WSN). Non-coherent combining is studied with average energy constraints. Because of the WSN setting, a weighted sum of the false alarm and the miss-detection probabilities is also combined in a single error term which is not suitable in cognitive networks as argued above. A soft sensing technique is considered in [31], where simultaneous transmission is used with maximum ratio combining (MRC) of the sensing results. This requires the knowledge of the reporting channel’s phase and magnitude at each local sensor. Furthermore, the authors solve for the achievable detection performance using a central limit theorem (CLT)
approximation.

In this chapter, we also consider the simultaneous transmission of the sensing results from all the cognitive radios (CRs) to the FC over fading reporting channels, and propose a novel cooperative sensing scheme, termed sensing with equal gain combining (SEGC). Under our model, each CR dephases its transmitted signal and equal gain combining of the sensors’ decisions is implemented at the FC. Relative to previous work, the contributions in this chapter are as follows:

i) We obtain the exact values of the global false alarm and miss-detection probabilities for the proposed SEGC technique by deriving the moment generating function (MGF) for the received signal at the FC. Our scheme can be efficiently implemented using the fast Fourier transform (FFT). While a Rayleigh fading distribution is assumed for the reporting channels between the CRs and the FC, this choice of distribution is not critical to the success of our scheme which can be extended to any general fading distribution, provided that the MGF for the received signal at the FC exists.

ii) In addition to quantifying the detection error probabilities, we also consider the secondary throughput as a performance metric. While detection error probabilities can be reduced by increasing the sensing time, this comes at the expense of reducing the transmission time, and hence the throughput, of the secondary user. This is the essence of the sensing-throughput tradeoff in cognitive radio networks [10].

iii) We demonstrate communication scenarios where equal gain combining of sensors’ decisions is superior to MRC, which is used in [31]. Our results are in harmony with other work in the literature, e.g. [32][33], which show that MRC receivers may be optimal in a data transmission diversity context, but not necessarily so in a decentralized detection setup. The proposed SEGC technique
also has a practical advantage over MRC since the transmitted waveform has a constant envelope. This allows radio power amplifiers to operate at maximum efficiency.

iv) We show the advantage of hard sensing over soft sensing under certain transmission conditions, specifically when the signal-to-noise ratio (SNR) is sufficiently high. Although previous work, e.g., shows a clear advantage for soft sensing over hard sensing in the presence of reporting channel errors, the result is confined to the orthogonal reporting of sensing results. To the best of our knowledge, no study exists on the difference between soft sensing and hard sensing when simultaneous reporting of sensors’ results is employed.

v) We also investigate the impact of realistic errors, such as phase and synchronization errors, and demonstrate the robustness of the proposed SEGC to such errors.

vi) We demonstrate that the Chernoff bound is not tight in our setup. Further, the CLT approximation, which is considered in, is shown not to be accurate when the local sensing time is small and when the number of sensors is small. Hence, an exact computation of the collaborative sensing performance is needed.

In the next section we detail the system model with a focus on the hard sensing case. In Section we study the impact of using soft sensing information and of having errors in phase estimation. Then, in Section we present some numerical results and, finally, we conclude the chapter in Section.

---

1It is important to note that EGC schemes are also considered in. However, therein the sensors’ decisions are reported in an orthogonal fashion and not simultaneously.

2Note that within the realm of orthogonal reporting of sensors’ decisions, the work in shows an advantage of hard sensing over soft sensing from the detection probabilities’ point of view, although the result is achieved using sequential detection and perfect reporting.
3.2 Collaborative Sensing With Equal Gain Combining

We consider a primary channel that operates in a time-slotted manner. Primary activity over the channel changes independently from one slot to another. A cognitive network of $N$ CRs coexists with the primary and tries to opportunistically access the primary channel when it is free. To realize this objective, the cognitive nodes carry out spectrum sensing at the beginning of each time slot, and then either make a binary decision on the state of primary activity and send it to the FC, called hard sensing, or directly transmit the received sensing power or energy to the FC, called soft sensing. In either case, a given cognitive radio (CR$_n$) transmits a decision/value $v_n$ to the FC. For the rest of this section, we focus on the hard sensing case. The soft sensing case is treated in Section 3.3.

Communication with the FC takes place over low-rate fading channels with limited bandwidth. The reporting channels function in a time division duplexing (TDD) fashion, which means that both directions of communication, from the radios to the FC and from the FC to the radios, operate using the same carrier frequency. The coherence time of the sensing and reporting channels is assumed to span many time slots. Channel estimation is performed periodically to track the channel variations. Specifically, the FC sends a pilot signal to all cognitive nodes. This pilot signal is used by each CR$_n$ ($n = 1, \cdots, N$) to estimate the phase introduced by the fading channel between itself and the FC. At the beginning of each time slot, each CR$_n$ performs spectrum sensing of the wireless medium to make a binary decision regarding the presence of the PU. All CRs then use ON/OFF keying (OOK) to simultaneously signal back to the FC the status of the PU. To allow for a coherent addition of the received signals at the FC, each CR$_n$ dephases its signal by multiplying it with the phase estimate obtained from the earlier FC pilot signal.
In this section, we assume that the phase estimates are perfect and we derive the performance metrics for the proposed SEGC scheme accordingly. We account for errors in the phase estimates in Section 3.3.2.

3.2.1 Statistics of the Received Signal at the FC

Based on the described model, we have the following received signal at the FC

\[ y = \sum_{n=1}^{N} w_n^* h_n v_n + Z, \]

(3.1)

where \( h_n = |h_n| e^{i\phi_n} \) (with magnitude \( |h_n| \) and phase \( \phi_n \)) is the complex channel gain between CR\(_n\) and the FC, and \( v_n \) is the local (hard) decision made by CR\(_n\) regarding the primary activity: \( v_n = 0 \) if the PU is sensed to be idle and \( v_n = 1 \) if it is sensed to be active.\(^3\) The term \( Z \sim \mathcal{N}(0, \sigma^2_S) \) is the receiver zero-mean Gaussian noise at the FC with variance \( \sigma^2_S \), parameter \( w_n \) is the weight factor used by the \( n^{th} \) sensor to multiply its decision, and \( (.)^* \) denotes the complex conjugation. For the SEGC scheme, we consider \( w_n = e^{i\phi_n} \). Thus, we have

\[ y = \sum_{n=1}^{N} |h_n| v_n + Z. \]

(3.2)

In the numerical analysis section, we also consider the MRC scheme, i.e. where \( w_n \) is proportional to \( h_n \), and compare its performance with the SEGC scheme. Note that the use of the word EGC herein should not be confused with its use in a diversity combining communication scheme, where the same symbol is sent over multiple antennas in the case of transmit diversity. In our distributed detection setup, possibly different symbols, representing the local decisions at the sensors, are sent over different reporting channels. In adopting the term EGC, we are referring to the fact that

\(^3\)Note that assuming a direct down-conversion in-phase/quadrature (IQ) receiver \([34]\), we use the received signal on the in-phase branch since there is no signal component on the quadrature branch in the case of perfect channel phase estimates.
the FC’s received signal in (3.2) is similar to the signal obtained via EGC in diversity combining techniques, hence the SEGC terminology.

We consider the null hypothesis, $H_0$, to refer to an idle PU. The alternative hypothesis, $H_1$, refers to an active PU. Under a threshold test on the received signal $y$, the false alarm and miss-detection probabilities are defined as

$$Q_{FA} = \Pr \{ y > \xi \mid H_0 \},$$  \hspace{1cm} (3.3)$$

$$Q_{MD} = \Pr \{ y < \xi \mid H_1 \},$$  \hspace{1cm} (3.4)$$

where $\xi$ is the global decision threshold at the FC. In order to calculate the above probabilities, the probability density functions (PDFs) $f_Y(y\mid H_0)$ and $f_Y(y\mid H_1)$ need to be found. Assuming independent sensor decisions and independent channel gains, the PDF of the received signal at the FC, conditioned on the sensors’ decisions $(v_1, v_2, \ldots, v_N)$, the channel gains $(h_1, h_2, \ldots, h_N)$ and either hypothesis, is Gaussian. That is,

$$f_Y \left( y \mid v_1, \ldots, v_N, h_1, \ldots, h_N, H_0 \right) = f_Y \left( y \mid v_1, \ldots, v_N, h_1, \ldots, h_N, H_1 \right)$$

$$= \frac{1}{\sqrt{2\pi}\sigma_S} \exp \left( -\frac{\left[ y - \sum_{n=1}^{N} |h_n|v_n \right]^2}{2\sigma_S^2} \right).$$  \hspace{1cm} (3.5)$$

The $v_n$’s are Bernoulli random variables. Thus, under the null hypothesis, we have

$$|h_n|v_n = \begin{cases} 
|h_n| & \text{w.p. } p_{FA,n} \\
0 & \text{w.p. } 1 - p_{FA,n}
\end{cases}$$  \hspace{1cm} (3.6)$$

where $p_{FA,n}$ is the local false alarm probability at CR$_n$, and where w.p. stands for ‘with probability’. Using (3.6) and by averaging over all the possible combinations of
the $v_1, v_2, \cdots, v_N$ values, we may write $f(y|h_1, ..., h_N, H_0)$ as

$$f_Y(y|h_1, ..., h_N, H_0) = \sum_{v_1, v_2, ..., v_N} \exp \left( -\frac{(y-\sum_{n=1}^{N} |h_n|v_n)^2}{2\sigma^2} \right) \frac{1}{\sqrt{2\pi\sigma^2}} \prod_{n=1}^{N} p_{FA,n}^{v_n} (1 - p_{FA,n})^{1-v_n}, \quad (3.7)$$

where $\sum_{v_1, v_2, ..., v_N}$ is a summation over all the possible values of the random binary sequence $v_1, v_2, \cdots, v_N$ starting from 000 to 111. Under hypothesis $H_1$, it is straightforward to show that

$$f_Y(y|h_1, ..., h_N, H_1) = \sum_{v_1, v_2, ..., v_N} \exp \left( -\frac{(y-\sum_{n=1}^{N} |h_n|v_n)^2}{2\sigma^2} \right) \frac{1}{\sqrt{2\pi\sigma^2}} \prod_{n=1}^{N} p_{D,n}^{v_n} (1 - p_{D,n})^{1-v_n}, \quad (3.8)$$

where $p_{D,n}$ is the local detection probability at CR$_n$.

In our model, we do not assume channel gain knowledge at the FC. This would require feeding back all the channel estimates to the FC, or making the cognitive nodes send pilots to the FC in a one-by-one fashion. In this case, we face the same problems that hinder orthogonal decision reporting such as the bandwidth requirements and the impact on throughput, albeit happening now only when the channel gains need to be updated. For this reason, we restrict channel knowledge at the FC to that of channel gain statistics. The PDFs $f_Y(y|H_0)$ and $f_Y(y|H_1)$ should be obtained from (3.7) and (3.8) via averaging over all the channel gains. However, there is no closed-form expression for $f_Y(y|H_0)$ and $f_Y(y|H_1)$ using (3.7) and (3.8), respectively, even when considering the simple case of two sensors simultaneously transmitting results. Hence, a different approach is needed which is discussed next.

### 3.2.2 The MGF Approach

Since the PDFs of $y$ conditioned on $H_0$ or $H_1$ are difficult to compute analytically, we resort to an efficient way to numerically compute the required probabilities in order to study the performance of the proposed SEGC scheme. The same approach is employed
in other works, e.g. \cite{37,40}, where the exact calculation of the detection probabilities is needed. We proceed by computing the MGF of \( y \) under the two hypotheses, and then we use the inverse discrete Fourier transform (IDFT) to compute the PDF. We detail the procedure for the MGF of \( y \) under the null hypothesis, denoted as \( \mathcal{M}_{Y|H_0}^{EGC,HS}(\rho|H_0) \)\footnote{Recall that we focus in this section on the hard sensing case.} and the corresponding MGF for \( y \) under the alternative hypothesis, denoted \( \mathcal{M}_{Y|H_1}^{EGC,HS}(\rho|H_1) \), follows in the same fashion.

By definition, we have

\[
\mathcal{M}_{Y|H_0}^{EGC,HS}(\rho|H_0) = \mathbb{E} \left[ \exp (\rho y) \mid H_0 \right],
\]

(3.9)

where \( \mathbb{E}(.) \) denotes the expectation operator. Using (3.2) in (3.9) and under the assumption of independence of all the involved random variables, the MGF can be evaluated as

\[
\mathcal{M}_{Y|H_0}^{EGC,HS}(\rho|H_0) = \mathbb{E} \left[ \exp (\rho Z) \right] \mathbb{E} \left[ \exp \left( \rho \sum_{n=1}^{N} |h_n| v_n \right) \mid H_0 \right] = \exp \left( \frac{1}{2} \rho^2 \sigma_S^2 \right) \prod_{n=1}^{N} \mathbb{E} \left[ \exp (\rho |h_n| v_n) \mid H_0 \right].
\]

(3.10)

Recall that the \( v_n \)'s are Bernoulli random variables so that \( v_n = 1 \) w.p. \( p_{FA,n} \), under the null hypothesis. We also assume a Rayleigh distribution for the random variables \( |h_n| \), i.e. the PDF of \( |h_n| \) is

\[
f(|h_n|) = \frac{|h_n|}{\varsigma_n^2} e^{-|h_n|^2/2\varsigma_n^2},
\]

(3.11)
information, we obtain,

\[
\mathcal{M}_Y^{\text{EGC,HS}}(\rho|\mathcal{H}_0) = \exp\left(\frac{1}{2} \rho^2 \sigma_x^2\right) \prod_{n=1}^{N} \left(1 + \sqrt{\pi} \frac{p_{\text{FA},n}s_n}{2} \exp \left(\frac{(\rho s_n)^2}{2}\right) \left[1 + \text{erf}\left(\frac{\rho s_n}{\sqrt{2}}\right)\right]\right),
\]

(3.12)

where \(\text{erf}(.)\) denotes the error function.

Proceeding in a similar way, we also obtain

\[
\mathcal{M}_Y^{\text{EGC,HS}}(\rho|\mathcal{H}_1) = \exp\left(\frac{1}{2} \rho^2 \sigma_x^2\right) \prod_{n=1}^{N} \left(1 + \sqrt{\pi} \frac{p_{\text{D},n}s_n}{2} \exp \left(\frac{(\rho s_n)^2}{2}\right) \left[1 + \text{erf}\left(\frac{\rho s_n}{\sqrt{2}}\right)\right]\right).
\]

(3.13)

Now, using the derived expressions for the MGF, the PDF of \(y\) can be evaluated by noting the following relationship:

\[
\mathcal{M}_Y^{\text{EGC,HS}}(-i2\pi f|\mathcal{H}_0) = \int_{-\infty}^{\infty} f_Y(y|\mathcal{H}_0) \exp(-i2\pi f y) \, dy.
\]

(3.14)

This means that \(\mathcal{M}_Y^{\text{EGC,HS}}(-i2\pi f|\mathcal{H}_0)\) and \(f_Y(y|\mathcal{H}_0)\) form a Fourier transform pair, and hence \(f_Y(y|\mathcal{H}_0)\) can be obtained from \(\mathcal{M}_Y^{\text{EGC,HS}}(-i2\pi f|\mathcal{H}_0)\) via the IDFT which can be efficiently implemented via variable discretization and use of the inverse fast Fourier transform (IFFT). This method is detailed in [41] and [42], together with an investigation of the accuracy and error bounds. Similarly, \(f_Y(y|\mathcal{H}_1)\) can be obtained from \(\mathcal{M}_Y^{\text{EGC,HS}}(-i2\pi f|\mathcal{H}_1)\). It is noted in (3.12) and (3.13) that an evaluation of the error function with complex arguments is needed. This can be achieved using a series representation of the error function as illustrated in [43].

In the performance analysis section, we will demonstrate the performance of the proposed hard sensing scheme and also contrast it with two modifications to the system model, which are discussed next.
3.3 SEGC Under Soft Sensing and Estimation Errors

In this section, we consider two different SEGC scenarios. In the first one, instead of each CR sending a binary decision regarding the presence of the PU as treated in the previous section, it instead sends the value of the measured sensing power, i.e., soft sensing is used. In the second scenario, we consider the case when phase errors occur so that the cognitive nodes cannot completely dephase their transmitted signals to the FC. Our goal in addressing these cases is to quantify the difference in performance in using hard versus soft information, and to also study the impact of practical errors, such as errors in phase estimation on the proposed collaborative sensing scheme.

3.3.1 Use of Soft Sensing Information

Herein, the setting is identical to the hard sensing model described in Section 3.2 except that the $v_n$’s now have the following distribution [44],

$$v_n = \begin{cases} 
\chi^2_{2u}, & \text{under } H_0 \\
\chi^2_{2u} (2\gamma_{s,n}), & \text{under } H_1 
\end{cases} \quad (3.15)$$

where $u$ is the time-bandwidth product and $\gamma_{s,n}$ is the sensing channel SNR between the PU and CR$_n$, $\chi^2_{2u}$ is a central Chi-squared random variable with $2u$ degrees of freedom, i.e.

$$f(v_n, 2u) = \frac{v_n^{u-1} e^{-v_n/2}}{2^u \Gamma(u)}, \quad (3.16)$$

where $\Gamma(a) = \int_0^\infty t^{a-1} \exp(-t) \, dt$ is the Gamma function. The variable $\chi^2_{2u} (2\gamma_{s,n})$ denotes a non-central Chi-squared random variable with $2u$ degrees of freedom and a
non-centrality parameter $2\gamma_s, n$, i.e.

$$f(v_n, 2u, 2\gamma_s, n) = \sum_{k=0}^{\infty} \frac{e^{-2\gamma_s, n\gamma_s}}{k!} f(v_n, 2(u + k)).$$  \hspace{1cm} (3.17)

Proceeding in a similar fashion to the analysis described in the previous section, the MGF under the null and alternative hypotheses can be written as

$$M_{Y_{EGC, SS}}(\rho|H_0) = \exp \left( \frac{\rho^2 \sigma_0^2}{2} \right) \prod_{n=1}^{N} \int_0^{\frac{1}{2\rho}} \frac{|h_n| \exp \left( -\frac{|h_n|^2}{2\gamma_s,n} \right)}{\sigma_n^2 (1 - 2\rho|h_n|)^u} d|h_n|,$$  \hspace{1cm} (3.18)

and

$$M_{Y_{EGC, SS}}(\rho|H_1) = \exp \left( \frac{\rho^2 \sigma_0^2}{2} \right) \prod_{n=1}^{N} \int_0^{\frac{1}{2\rho}} \frac{|h_n| \exp \left( \frac{2\gamma_s,n\rho|h_n|}{1-2\rho|h_n|} \right) \exp \left( -\frac{|h_n|^2}{2\gamma_s,n} \right)}{\sigma_n^2 (1 - 2\rho|h_n|)^u} d|h_n|,$$  \hspace{1cm} (3.19)

respectively.

The above integrals can be evaluated numerically and the corresponding PDFs evaluated for the SEGC with soft sensing by making use of the relationship in (3.14). It is noted that the expression in (3.19) assumes that the $\gamma_s, n$’s are known at the FC. This assumption provides the best performance possible for the soft sensing scheme and, hence, serves as an upperbound. In the performance analysis section, we compare the performance of the soft and hard sensing and show that there are cases where hard sensing is superior to soft sensing, particularly when the SNR is sufficiently high. Before that, we address the incorporation of practical errors, such as phase errors, in the SEGC analysis.
3.3.2 Errors in Phase Estimation

We begin by analyzing the signals at the sensing nodes which are used for phase estimation. Recall that the FC sends a pilot signal to all CRs so that phase information on the channels between the CRs and the FC can be acquired. Using a typical in-phase/quadrature (I/Q) receiver, the received in-phase and quadrature components at a given CR\(n\) can be written as,

\[
Y_{I,n} = \sqrt{P_{FC}} |h_n| \cos \varphi_n + Z_{I,n}, \tag{3.20}
\]

\[
Y_{Q,n} = \sqrt{P_{FC}} |h_n| \sin \varphi_n + Z_{Q,n}, \tag{3.21}
\]

where \(P_{FC}\) is the power used by the FC to send the pilot signal, \(\varphi_n\) is the actual phase of the channel between the FC and CR\(n\), and where \(Z_{I,n}\) and \(Z_{Q,n}\) are the in-phase and quadrature receiver noise, each distributed according to \(N(0, \sigma^2_{\delta,n})\).

We consider a maximum likelihood (ML) phase estimator at each CR. It can be verified in a straightforward manner that the ML estimator in this case is given by

\[
\hat{\varphi}_n = \tan^{-1} \left( \frac{Y_{Q,n}}{Y_{I,n}} \right). \tag{3.22}
\]

Now, using an I/Q receiver at the FC, the received signal will be,

\[
y = \sum_{n=1}^{N} |h_n| v_n \cos(\delta \varphi_n) + Z, \tag{3.23}
\]

where \(\delta \varphi_n = \varphi_n - \hat{\varphi}_n\) is the difference between the true channel phase and the estimated one, pertaining to the \(n\)th CR.

---

\(^5\)The analysis here is restricted to the hard sensing case to highlight the impact of phase errors. Extension to the soft sensing scenario is straightforward.

\(^6\)In case more than one pilot symbol is used for phase estimation, the CR can average the in-phase and quadrature components and apply (3.22) to the averages. In this case, the noise variance becomes \(\sigma^2_{\delta,n}\) divided by the number of pilot symbols.
Proceeding in a similar fashion to the previous section, we derive the modified MGF under the two hypotheses, now denoted $\hat{M}_{Y}^{\text{EGC,HS}}(\rho|\mathcal{H}_0)$ and $\hat{M}_{Y}^{\text{EGC,HS}}(\rho|\mathcal{H}_1)$, to be,

$$\hat{M}_{Y}^{\text{EGC,HS}}(\rho|\mathcal{H}_i) = \exp\left(\frac{1}{2}\rho^2\sigma_n^2\right) \prod_{n=1}^{N} \mathbb{E}\left[\exp\left(\rho|h_n|v_n\cos(\delta\phi_n)\right)|\mathcal{H}_i\right],$$

(3.24)

where $i \in \{0, 1\}$. Because the MGF calculation now involves an averaging over the Bernoulli distributed $v_n$’s, the Rayleigh random variables $|h_n|$’s and the $\delta\phi_n$’s, the expressions in (3.24) become quite lengthy and complicated, hence we relegate their reporting to Appendix 3.6.1.

Next, we illustrate the performance of the proposed SEGC for hard sensing, soft sensing and phase errors.

### 3.4 Performance Analysis of SEGC

In this section, we start by solving an optimization problem to derive the local and global thresholds. We then introduce a TDMA system which employs orthogonal reporting of sensing results to compare with the SEGC scheme in terms of throughput and detection performance. We also compare the performance of soft and hard sensing under the SEGC scheme and also investigate the case where MRC is employed instead of EGC. Finally, we test the robustness of the SEGC scheme against phase and synchronization errors and assess the SEGC’s performance against variations in the average SNR and the reporting SNR.

#### 3.4.1 Optimizing the Local and Global Thresholds

Here, we focus on the hard sensing scenario. Using the results of [44] and assuming Rayleigh fading for the sensing channels at all CRs, the local false-alarm and detection
probabilities are given by
\[ p_{FA,n} = \Gamma_{\text{inc}}(u, \xi_n/2), \]  
(3.25)

and
\[ p_{D,n} = e^{-\frac{\xi_n}{2}} \sum_{l=0}^{u-2} \frac{1}{l!} \left( \frac{\xi_n}{2} \right)^l + \left( 1 + \frac{\gamma_{s,n}}{\varphi_{s,n}} \right)^{u-1} \left[ e^{-\frac{\xi_n}{2(1+\gamma_{s,n})}} - e^{-\frac{\xi_n}{2}} \sum_{l=0}^{u-2} \frac{1}{l!} \left( \frac{\xi_n}{2(1+\gamma_{s,n})} \right)^l \right], \]  
(3.26)

respectively, where \( \Gamma_{\text{inc}}(.,.) \) is the normalized incomplete Gamma function given by
\[ \Gamma_{\text{inc}}(x,a) = \frac{1}{\Gamma(a)} \int_0^x t^{a-1} \exp(-t) \, dt, \]  
(3.27)

the quantity \( u \) is chosen to have an integer value of 5 unless otherwise stated, \( \xi_n \) is the local decision threshold at CR\(_n\) and \( \gamma_{s,n} \) is the average SNR of the sensing channel between the PU and CR\(_n\).

In order to calculate the local and global thresholds, we consider the case when the primary network specifies a maximum miss-detection threshold, \( Q^{(T)}_{MD} \), that cannot be exceeded. The FC solves the following Neyman-Pearson-like optimization problem:

\[ \min_{\xi, \xi_1, \xi_2, \ldots, \xi_N} Q_{FA} \]

subject to: \( Q_{MD} \leq Q^{(T)}_{MD} \),
(3.28)

where \( \xi \) is the global threshold at the FC as indicated in expressions (3.3) and (3.4).

The problem is non-convex and can be solved numerically via gradient descent techniques for every coherence time of the channels.

Fig. 3.1 shows the resulting global false-alarm curve, \( Q_{FA} \) given by (3.3), against the specified miss-detection probability, \( Q^{(T)}_{MD} \), for \( N = 20 \) sensors. The average sensing SNR is 5dB at all sensors, and the reporting SNR to the FC is also 5dB. The \( \varsigma_n \)'s used in this figure are calculated to be the square root of the reporting SNR multiplied by the noise variance, i.e. \( \varsigma_n = \sqrt{\bar{\gamma}_{r,n} \sigma^2_S} \), where \( \bar{\gamma}_{r,n} \) is the average reporting
Figure 3.1: Complementary receiver operation characteristics (C-ROC) curves for \( N = 20 \) sensors.

SNR for CR\(_n\). Also shown on the figure is another curve obtained via a Monte Carlo system simulation over a large number of time slots. The results clearly validate the computation of the PDF of \( y \) from the MGF using the IFFT. Furthermore, comparing Matlab run times, we found the computational time for system simulation averages around 0.35 time units while that of the SEGC scheme averages around 0.003 time units. This means our scheme is about two orders of magnitude faster.

We also compute the Chernoff bound for the SEGC, which can be shown to be,

\[
Q_{FA} \leq \min_{\rho > 0} \exp (-\rho \xi) \exp \left( \frac{1}{2} \rho^2 \sigma_S^2 \right) \prod_{n=1}^{N} \left( 1 + \sqrt{2\pi} p_{FA,n} \varsigma n \rho \exp \left( \frac{(\rho \varsigma_n)^2}{2} \right) [1 - Q(\rho \varsigma_n)] \right),
\]

\[
Q_{MD} \leq \min_{\rho > 0} \exp (\rho \xi) \exp \left( \frac{1}{2} \rho^2 \sigma_S^2 \right) \prod_{n=1}^{N} \left( 1 - \sqrt{2\pi} p_{D,n} \varsigma n \rho \exp \left( \frac{(\rho \varsigma_n)^2}{2} \right) Q(\rho \varsigma_n) \right),
\]

where \( Q(x) = \frac{1}{\sqrt{2\pi}} \int_{x}^{\infty} \exp \left( -\frac{t^2}{2} \right) dt \) is the tail probability of the standard normal distribution. Also superimposed on this figure, is a computation of the CLT approximation, used in [31], which can be shown to yield the following global detection probabilities\(^7\)

\[
Q_{FA,CLT} = Q \left( \frac{\xi - \mu_0}{\sigma_0} \right),
\]

\(^7\)For simplicity, here we assume identical local detection probabilities at the cognitive nodes so that \( p_{FA,n} = p_f \) and \( p_{D,n} = p_d \). We also assume identical reporting channels so that \( \varsigma_n = \varsigma \).
\[ Q_{\text{MD, CLT}} = Q \left( \frac{\mu_1 - \xi}{\sigma_1} \right), \]  

where \( \mu_0 = \sqrt{\frac{\pi}{2}} N \Delta \), \( \mu_1 = \sqrt{\frac{\pi}{2}} N \Delta \), \( \sigma_0^2 = N \Delta p_{\text{FA}} (2 - \pi p_{\text{FA}} / 2) + \sigma_S^2 \) and \( \sigma_1^2 = N \Delta p \Delta (2 - \pi p_{\text{FA}} / 2) + \sigma_S^2 \). It is clear from the figure that the Chernoff bound is not tight. Also, the CLT approximation is not accurate despite the large number of sensors and the large \( u \) value.

Next, we compare the SEGC scheme with orthogonal reporting of the sensors’ decisions.

### 3.4.2 A TDMA System and Throughput Comparison

We now consider a TDMA scheme where the FC has access to a noisy version of each sensor’s decision. For a fair comparison between the proposed SEGC scheme and this TDMA system, we propose the following collaborative sensing system:

1. Channel phase estimation takes place in a way that is identical to SEGC. That is, the FC sends a pilot signal to all CRs, which is used to estimate the channel phase between each CR and the FC.

2. Each CR sends its dephased decision on a designated mini-slot within the slot duration, according to an agreed-upon schedule with the FC. This means that the received signal at the FC from CR\(_n\) in the \( n \)th mini-slot will be,

\[ y_n = |h_n| v_n + Z_n, \]  

where \( Z_n \sim \mathcal{N}(0, \sigma_S^2) \).

3. The FC now makes its global decision by calculating the following log-likelihood

\[ \text{The analysis is carried out here for the hard sensing case. Extension to the soft sensing case is straightforward.} \]
ratio (LLR),

\[ \text{LLR}_{\text{TDMA}} = \sum_{n=1}^{N} \log \left( \frac{f(y_n|H_0)}{f(y_n|H_1)} \right) \overset{H_0}{\gtrless} \xi, \quad (3.34) \]

where the PDFs \( f(y_n|H_0) \) and \( f(y_n|H_1) \) can be attained in closed-form (cf. Appendix 3.6.2.). The global false-alarm and miss-detection probabilities are given by \( \Pr\{\text{LLR}_{\text{TDMA}} < \xi|H_0\} \) and \( \Pr\{\text{LLR}_{\text{TDMA}} > \xi|H_1\} \), respectively. Note that this use of the LLR is the optimal fusion of the various observations. Various suboptimal approaches exist such as, for instance, decoding each binary decision \( v_n \) using \( y_n \) and then fusing the decoded decisions using the OR, AND, or \( K \)-out-of-\( N \) rules. Our choice here gives the best possible performance for the TDMA system.

In addition to computing the detection error probabilities, we also use throughput as a performance metric that is more relevant in the context of CR. It is possible under a TDMA scheme to wait for a large number of observations and reduce the detection errors to very small values. Nevertheless, the PU may change its state of activity over a relatively short period, thereby making it necessary for the CR to make a fast, yet reliable, decision in order to obtain a transmission opportunity. This requirement creates a sensing reliability-throughput tradeoff that is characteristic of CR operation \[10\].

Under the SEGC scheme, we define throughput as

\[ R_{S,\text{SEGC}} = (1 - Q_{\text{FA}}) \left( \frac{T - \tau - \tau_s}{T} \right) r_S, \quad (3.35) \]

where \( Q_{\text{FA}} \) is given by (3.3), \( T \) is the time slot duration, \( \tau \) is the time taken by the CRs to send their sensing decisions to the FC, \( \tau_s \) is the fraction of time used by the CRs to carry out local sensing and \( r_S \) is the mean transmission rate achieved by the secondary node that seizes the transmission opportunity.\footnote{We assume here that once a vacancy is observed, one CR seizes the transmission opportunity.}
Figure 3.2: C-ROC curves for $N = 30$ sensors for the SEGC and TDMA schemes at reporting SNRs of $5\text{dB}$ and $17\text{dB}$.

that miss-detection leads to the loss of the secondary packet and, hence, no rate is achieved.

For the TDMA scheme, on the other hand, we have

$$R_{S,\text{TDMA}} = (1 - Q_{\text{FA,TDMA}}) \left( \frac{T - N \tau - \tau_s}{T} \right)^+ r_s, \quad (3.36)$$

where $(x)^+ = \max\{x, 0\}$ and $Q_{\text{FA,TDMA}}$ is the global false alarm probability for the TDMA scheme, calculated numerically as the $\Pr\{\text{LLR}_{\text{TDMA}} < \xi | H_0\}$. Note that, as expected, the throughput of the TDMA scheme could drop to zero as the number of sensors, $N$, increases.

Given the above equations, our throughput-based optimization problem is formulated as follows

$$\max_{\xi, \xi_1, \xi_2, \ldots, \xi_N} R$$

subject to: $Q_{\text{MD}} \leq Q_{\text{MD}}^{(T)}$, \quad (3.37)

where $R$ is the throughput given by $\eqref{eq:segc2}$ and $\eqref{eq:tdma2}$ for the SEGC and TDMA schemes, respectively.
Fig. 3.2 shows the solution of the problem (5.10) for the SEGC scheme and the outlined TDMA system for the case of \( N = 30 \) sensors. The average sensing SNR is set to the same value for both schemes (5dB) so that we can evaluate the difference in their performances over the sensor reporting part only. The PDF of the LLR_{TDMA} is computed numerically using \( y_n \)'s generated from the PDFs \( f(y_n | H_0) \) and \( f(y_n | H_1) \). As observed, the performance of both systems degrades with the decrease in reporting SNR value. The degradation is negligible in the case of SEGC, which shows the robustness of the system towards variations in the reporting SNR. The performance of the TDMA system, however, varies dramatically with the reporting SNR. In the low-SNR regime, where each TDMA reading suffers from noise albeit being decoded separately, SEGC is superior. The TDMA scheme does not improve significantly above the SEGC except in the high-SNR case.

Also superimposed on the curve is the resulting ROC for an upperbound system which we construct through the use of [37]. In that work, orthogonal reporting of sensors’ decisions is employed. However, the reporting is perfect, i.e. no channel fading or AWGN exists. It is not difficult to see in this case why the resulting ROC upperbounds the proposed SEGC scheme. To setup the comparison, we use the same LLR_{TDMA} metric as in (3.34). However, in this case the PDFs \( f(y_n | H_0) \) and \( f(y_n | H_1) \) are readily attainable as

\[
f(y_n | H_0) = p_{FA,n}^{v_n} (1 - p_{FA,n})^{1-v_n}, \tag{3.38}
\]

\[
f(y_n | H_1) = p_{D,n}^{v_n} (1 - p_{D,n})^{1-v_n}. \tag{3.39}
\]

It is noted that the gap is wide between the proposed solution and the upperbound. However, Fig. 3.3 shows the achievable throughput by solving (5.16) for the SEGC and the TDMA schemes, in addition to the upperbound system. The superiority of the SEGC scheme to the upperbound and the other TDMA schemes is clear and is
mainly because of the short reporting overhead. Things improve a bit for the TDMA system with the increase in the reporting SNR. However, the achievable throughput remains far from what is possible using the SEGC. For this figure, we chose \( \tau = 0.01T \) while \( \tau_s = 0.005T \).\(^{10}\)

Having compared the SEGC with orthogonal reporting techniques, we now proceed to making a comparison with another technique which also uses simultaneous transmission of sensors’ decisions.

### 3.4.3 Comparing MRC and SEGC

Herein, we investigate whether the cooperative sensing performance can be improved over SEGC by making available both channel phase and magnitude at each CR, in analogy to the gains achieved by wireless communication diversity techniques employing MRC, which requires knowledge of channel gain and magnitude.

When MRC is used, the FC’s received signal is given by,

\[
y = \sum_{n=1}^{N} \frac{|h_n|^2 v_n}{\sqrt{2\kappa_n^2}} + Z, \tag{3.40}
\]

\(^{10}\)The value of \( \tau \) is selected to be of the same order of magnitude as the suggested values in \(^{45}\), while \( \tau_s \) is set to a reasonable value given the sensing SNR of 5dB and the sensing-throughput tradeoff as highlighted in \(^{10}\).
i.e. the weight factor is \( w_n = h_n / \sqrt{2 \varsigma_n^2} \). Based on (3.1), we note that the average transmit power of each local sensor will be \( \mathbb{E}(|w_n|^2v_n^2) = \mathbb{E}(|w_n|^2)\mathbb{E}(v_n^2) \). The normalization by the channel’s \( \varsigma_n \) value in (3.40) above ensures that MRC will have the same average transmit power as the corresponding SEGC, for fair comparison.

By doing a similar analysis to the SEGC scheme and noting that now the \(|h_n|^2\)'s are exponentially distributed random variables with parameter \( \varsigma_n \), we may write the MGF of the MRC scheme as follows

\[
M_{Y,\text{MRC, HS}}(\rho|\mathcal{H}_0) = \exp \left( \frac{1}{2} \rho^2 \sigma_S^2 \right) \prod_{n=1}^{N} \left( 1 - p_{\text{FA},n} + p_{\text{FA},n} \left( 1 - \frac{\varsigma_n \rho}{\sqrt{2}} \right)^{-1} \right),
\]

(3.41)

\[
M_{Y,\text{MRC, HS}}(\rho|\mathcal{H}_1) = \exp \left( \frac{1}{2} \rho^2 \sigma_S^2 \right) \prod_{n=1}^{N} \left( 1 - p_{\text{D},n} + p_{\text{D},n} \left( 1 - \frac{\varsigma_n \rho}{\sqrt{2}} \right)^{-1} \right).
\]

(3.42)

We can then derive the PDFs using the IFFT method, as was previously done throughout the paper.

Figure 3.4: Comparison between SEGC and MRC of sensors’ decisions for \( N = 10 \) sensors under both soft-sensing and hard-sensing. The sensing SNR is set to 10dB, the reporting SNR is \(-10\)dB and \( u = 1 \). SEGC with hard-sensing gives the best performance.

Interestingly enough, we find that SEGC does perform better than MRC in many
Figure 3.5: Comparison between SEGC and MRC of sensors’ decisions for \( N = 40 \) sensors under both soft-sensing and hard-sensing. The sensing and reporting SNRs are set to 0dB and \( u = 1 \). SEGC with soft sensing provides the best performance.

cases. Figures 3.4 and 3.5 show the achievable C-ROC, i.e. \( Q_{FA} \) versus \( Q_{MD} \) curve, of the proposed SEGC in comparison to the MRC scheme for both hard and soft sensing under variation of the different system parameters. Specifically, we highlight cases where hard sensing is better than soft sensing and where soft sensing is better than hard sensing. In both cases, the SEGC scheme is superior to the MRC one. Previous work, e.g. \[32,33\], also report cases where EGC is better than MRC, in an orthogonal reporting setting. In Appendix 3.6.3., we provide a more detailed discussion on this topic.

3.4.4 Impact of Phase and Synchronization Errors

Figure 3.6 shows the achievable performance under the consideration of errors in the phase estimate. Also imposed on this figure is the system simulation results with phase errors. The loss in performance is small and improves by increasing the value of \( P_{FC} \), which is the power of the pilot symbol transmitted by the FC.

In Fig. 3.7, we plot the SEGC performance when both synchronization and phase errors occur. To calculate the synchronization error, we consider the transmission of rectangular pulses of duration \( \tau \) by each CR. This means that the received signal now
Figure 3.6: SEGC performance under phase errors, which cause an increase in $Q_{MD}$. The degradation is reduced by increasing the power of pilot symbols, $P_{FC}$. Also shown are curves obtained via system simulation for validation purposes.

Figure 3.7: SEGC performance under phase and synchronization errors for $P_{FC} = 1$. 
where \( \tau \) is the time interval during which the sensors transmit their decisions to the FC and \( \iota_n \) is the timing error between \( \text{CR}_n \) and the FC. The reported results are for uniformly distributed \( \iota_n \)'s from \([-0.1\tau, 0.1\tau]\) and also for the case of \( \iota_n \)'s uniformly distributed in \([-0.5\tau, 0.5\tau]\). The results clearly demonstrate the robustness of SEGC to synchronization errors, which can be explained by the fact that such errors lead to reduction in the effective SNR at the FC. SEGC, however, is robust to variations in the reporting SNR as demonstrated by Fig. 3.8 which shows the variation of the minimum \( Q_{\text{FA}} \) with the SNR, for both the sensing SNR and the reporting SNR. It is clear that performance improves significantly when the sensing SNR improves. On the other hand, fluctuations in the reporting SNR have a relatively weak impact on performance, further validating the robustness of our SEGC scheme to variations in the reporting SNR.

![Figure 3.8: Variation of the minimum \( Q_{\text{FA}} \) with the SNR.](image)

### 3.5 Conclusion

We proposed a collaborative sensing scheme for cognitive radio networks termed sensing with equal gain combining (SEGC). Under the SEGC scheme, sensor nodes si-
multaneously transmit their sensing decisions to be coherently combined at the fusion center. We obtained the global detections probabilities exactly through an MGF approach. We considered both cases of hard and soft sensing and provided communication scenarios where hard sensing is superior to its soft counterpart. Contrary to the schemes implementing orthogonal reporting of sensors’ decisions, our findings document instances where hard sensing is superior to soft sensing. We also compared our approach with the MRC scheme and showed examples where the SEGC method, bearing only the cost of channel phase information, is superior. From the reliability of detection perspective, the proposed approach outperforms a comparison TDMA system at low SNRs. From a secondary throughput perspective, our scheme always outperforms TDMA and other orthogonal decision reporting techniques where throughput suffers with the increase in the number of sensors. We also characterized our system performance from the perspective of variation of the sensing and reporting SNRs and with consideration to phase and synchronization errors as encountered in practice. We showed the robustness of our scheme under such circumstances.

3.6 Appendix

3.6.1 Evaluation of MGFs under Errors in Phase Estimation

To calculate (3.24), we need to evaluate the term $\mathbb{E}(\exp (\rho |h_n| v_n \cos(\delta \varphi_n)) | \mathcal{H}_i)$, $i \in \{0, 1\}$, for every sensing node $n$. Herein, we drop the index dependency for notational convenience and define $g = |h|$. We also drop the conditioning on the hypotheses. Thus, we may write,

$$\mathbb{E}(\exp (\rho g v \cos(\delta \varphi))) = \sum_{v=0,1} \int_0^{2\pi} \int_0^{\infty} \exp(\rho vg \cos(\delta \varphi)) f(\delta \varphi | g) f(g) dg d\delta \varphi p(v).$$

(3.44)
Then averaging over the decisions $v$, we get

$$
\mathbb{E}(\exp(\rho gv \cos(\delta \varphi))) = p(v = 0) + p(v = 1) \int_0^{2\pi} \int_0^\infty \exp(\rho g \cos(\delta \varphi)) f(\delta \varphi|g) f(g) dg d\delta \varphi.
$$

(3.45)

where $p(v = 0)$ is obtained as,

$$
p(v = 0) = \begin{cases} 
1 - p_{FA} & \text{under } \mathcal{H}_0 \\
1 - p_D & \text{under } \mathcal{H}_1,
\end{cases}
$$

(3.46)

while $p(v = 1)$ is calculated as,

$$
p(v = 1) = \begin{cases} 
p_{FA} & \text{under } \mathcal{H}_0 \\
p_D & \text{under } \mathcal{H}_1.
\end{cases}
$$

(3.47)

To evaluate (3.45), we need to derive the PDF of the estimated phase error, conditioned on a particular channel realization, $f(\delta \varphi|g)$. Using (3.20) and (3.21), define an auxiliary random variable $Z$ such that

$$
Z = \sqrt{Y_I^2 + Y_Q^2}.
$$

(3.48)

By a simple transformation of random variables from $Y_I$ and $Y_Q$ to $Z$ and $\delta \varphi$, the joint PDF of $Z$ and $\delta \varphi$, $f(Z, \delta \varphi|g)$, may be attained as

$$
f(Z, \delta \varphi|g) = \frac{Z}{2\pi \sigma_Z^2} \exp\left(-\frac{Z^2 + P_{FC}g^2 - 2Z\sqrt{P_{FC}g} \cos(\delta \varphi)}{2\sigma_Z^2}\right).
$$

(3.49)
Integrating over $Z$, we get

$$f(\delta \varphi | g) = \frac{1}{2\pi} \exp\left(-\eta g^2\right) + \sqrt{\frac{\eta}{\pi}} g \cos(\delta \varphi) \exp\left(-\eta g^2 + \eta g^2 \cos^2(\delta \varphi)\right) - \sqrt{\frac{\eta}{\pi}} g \cos(\delta \varphi) \exp\left(-\eta g^2\right) \exp\left(\eta g^2 \cos^2(\delta \varphi)\right) Q \left(\cos(\delta \varphi) \sqrt{2\gamma g^2}\right),$$  \hspace{1cm} (3.50)

where $\eta = \frac{P_{FC}}{2\sigma^2}$. Next, we evaluate the integral $\int_{0}^{\infty} \exp(\rho g \cos(\delta \varphi)) f(\delta \varphi | g) f(g) dg$ in (3.45). From (3.50), there are three expressions to evaluate and we take each separately. For the first expression, we have

$$\int_{0}^{\infty} \exp\left(-\eta g^2\right) \exp(\rho g \cos(\delta \varphi)) \frac{g}{2\pi \varsigma^2} \exp\left(-\frac{g}{2\varsigma^2}\right) dg = \exp\left(\frac{\rho^2 \cos^2(\delta \varphi)}{4w}\right) \int_{\rho^2 \cos^2(\delta \varphi)}^{\infty} \exp\left(-\beta\right) \left[\frac{\rho \cos(\delta \varphi)}{2w} + \beta^{1/2}\right] \frac{d\beta}{2\pi \varsigma^2 2\sqrt{\beta w}},$$  \hspace{1cm} (3.51)

where $\beta = \left[g - \frac{\rho \cos(\delta \varphi)}{2w}\right]^2$, $w = \eta + \frac{1}{2\varsigma^2}$ and we arrive at the equation above by a change of integration variables. Continuing further, we get

$$\exp\left(\frac{\rho^2 \cos^2(\delta \varphi)}{4w}\right) \int_{\rho^2 \cos^2(\delta \varphi)}^{\infty} \exp\left(-\beta\right) \left[\frac{\rho \cos(\delta \varphi)}{2w} + \beta^{1/2}\right] \frac{d\beta}{2\pi \varsigma^2 2\sqrt{\beta w}} = \frac{1}{4\pi w \varsigma^2} + \frac{\rho \cos(\delta \varphi) \exp\left(\frac{\rho \cos(\delta \varphi)}{4w}\right)}{8\sqrt{\pi} \varsigma^2 w^{3/2}} \left[1 - \Gamma_{\text{inc}}\left(\frac{\rho^2 \cos^2(\delta \varphi)}{4w}, \frac{1}{2}\right)\right].$$  \hspace{1cm} (3.52)

Proceeding in the same fashion for the second expression, we obtain (3.53),

$$\frac{\sqrt{\eta} \cos(\delta \varphi)}{2\pi^{3/2} \varsigma^2} \int_{0}^{\infty} g^2 \exp\left(-\eta g^2\right) \exp(\eta g^2 \cos^2(\delta \varphi)) \exp\left(-\frac{g^2}{2\varsigma^2}\right) \exp(\rho g \cos(\delta \varphi)) dg =$$

$$\frac{\sqrt{\eta} \cos(\delta \varphi)}{4\pi^{3/2} \varsigma^2 \sqrt{x}} \left[\frac{\rho \cos(\delta \varphi)}{x^{3/2}} + \exp\left(\frac{\rho^2 \cos^2(\delta \varphi)}{4x}\right) \frac{\rho^2 \cos^2(\delta \varphi)}{4x^2} \sqrt{\pi} \left(1 - \Gamma_{\text{inc}}\left(\frac{\rho^2 \cos^2(\delta \varphi)}{4x}, 1/2\right)\right)\right]$$

$$+ \exp\left(\frac{\rho^2 \cos^2(\delta \varphi)}{4x}\right) \frac{\sqrt{\pi}}{2x} \left[1 - \Gamma_{\text{inc}}\left(\frac{\rho^2 \cos^2(\delta \varphi)}{4x}, 3/2\right)\right],$$  \hspace{1cm} (3.53)

where $x = \eta \sin^2(\delta \varphi) + \frac{1}{2\varsigma^2}$. The third expression, given by
is evaluated numerically using Matlab. Finally, all the attained terms, \((3.52), (3.53)\) and \((3.54)\), need to be averaged out with respect to \(\delta \varphi\), which we also implement in Matlab. Having obtained the MGFs, we can proceed to obtain the PDFs via the IFFT procedure and then compute the detection error probabilities.

### 3.6.2 The PDFs under the TDMA Scheme

Using equation \((3.33)\), we write

\[
f(y_n|h_n, \mathcal{H}_o) = \frac{1-p_{FA,n}}{\sqrt{2}\pi \sigma_S} \exp\left( -\frac{y_n^2}{2\sigma_S^2} \right) + \frac{p_{FA,n}}{\sqrt{2}\pi \sigma_S} \exp\left( -\frac{(y_n - |h_n|)^2}{2\sigma_S^2} \right),
\]

(3.55)

and similarly,

\[
f(y_n|h_n, \mathcal{H}_1) = \frac{1-p_{D,n}}{\sqrt{2}\pi \sigma_S} \exp\left( -\frac{y_n^2}{2\sigma_S^2} \right) + \frac{p_{D,n}}{\sqrt{2}\pi \sigma_S} \exp\left( -\frac{(y_n - |h_n|)^2}{2\sigma_S^2} \right).
\]

(3.56)

By averaging out the channel gains, we arrive at

\[
f(y_n|\mathcal{H}_o) = \frac{1-p_{FA,n}}{\sqrt{2}\pi \sigma_S} \exp\left( -\frac{y_n^2}{2\sigma_S^2} \right) + \frac{p_{FA,n}}{\sqrt{2}\pi \sigma_S} A,
\]

(3.57)

and

\[
f(y_n|\mathcal{H}_1) = \frac{1-p_{D,n}}{\sqrt{2}\pi \sigma_S} \exp\left( -\frac{y_n^2}{2\sigma_S^2} \right) + \frac{p_{D,n}}{\sqrt{2}\pi \sigma_S} A,
\]

(3.58)

respectively, where \(A\) can be evaluated to be

\[
A = \frac{1}{2\zeta_n} \exp\left( -\frac{y_n^2}{4\sigma_S^2\zeta_n} \right) + \sqrt{\frac{\pi}{\zeta_n}} \left( \frac{y_n}{2\sigma_S^2\zeta_n} \right) \exp\left( -\frac{y_n^2\zeta_n}{\zeta_n^2} \right) \left( 1 - Q \left( \frac{y_n}{\sigma_S^2\sqrt{2\zeta_n}} \right) \right),
\]

(3.59)

with \(\zeta_n = \frac{1}{2} \left( \frac{1}{\zeta_n} + \frac{1}{\sigma_S^2} \right)\).
3.6.3 MRC versus EGC of Sensors’ Decisions

In this appendix, we briefly discuss the selection of the weights \( w_n \) in (3.1). In a transmit diversity context, with known channel state information, the received signal is

\[
y = \sum_{n=1}^{N} w_n^* h_n s + Z,
\]

where symbol \( s \) is transmitted from all the antennas. In that case, it can be shown that \( w_n \propto h_n \), i.e. MRC, maximizes the received SNR and, hence, is optimal in this setting [46]. On the other hand, in the distributed detection scenario considered in this paper, possibly different symbols are transmitted by the different sensors and a question arises as to what could be the optimal weighting vector if we extend the optimization problem in (5.10) to include the weights. This extension would significantly increase the complexity of the problem. So, in order to gain insight into the proper choice of \( w_n \) and to explain our results which show that MRC is not necessarily optimal in a distributed detection setting, we consider a simpler optimization problem that instead considers the deflection coefficient as an objective.

Specifically, we aim to maximize the modified deflection coefficient (MDC) [47], given by

\[
d^2 = \frac{|\mathbb{E}(y|\mathcal{H}_1) - \mathbb{E}(y|\mathcal{H}_0)|^2}{\text{var}(y|\mathcal{H}_1)},
\]

where \( \text{var} \) denotes the variance of a random variable. Given (3.1), it can be easily shown that

\[
\mathbb{E}(y|\mathcal{H}_0) = \sum_{n=1}^{N} w_n^* h_n p_{FA,n}, \quad (3.61)
\]

\[
\mathbb{E}(y|\mathcal{H}_1) = \sum_{n=1}^{N} w_n^* h_n p_{D,n}, \quad (3.62)
\]

and

\[
\text{var}(y|\mathcal{H}_1) = \sigma_S^2 + \sum_{n=1}^{N} |w_n|^2 |h_n|^2 p_{D,n} (1 - p_{D,n}). \quad (3.63)
\]
Note that
\[ \mathbb{E}(y|H_1) - \mathbb{E}(y|H_0) = w^*g, \tag{3.64} \]
where \( g = [h_1(p_{d,1} - p_{f,1}) ... h_N(p_{D,n} - p_{F_A,n})]^T \), \( w = [w_1, w_2, ..., w_N] \) is the vector of sensor weights and we use \((_.)^T\) to denote the transpose operation. Defined \( m = [(|h_1|^2 p_{d,1} (1 - p_{d,1}) ... |h_N|^2 p_{D,n} (1 - p_{D,n})]^T \) and \( M \) as the \( N \times N \) diagonal matrix with the elements of \( m \) on the main diagonal, then the MDC as defined in (3.60) can be equivalently written as
\[ d^2 = \frac{|w^*g|^2}{\sigma_S^2 + w^*Mw}. \tag{3.65} \]
Applying the Cauchy-Schwarz inequality, it can then be shown that \( d^2 \) is maximized by choosing the weights according to:
\[ w_n \propto \frac{h_n [p_{D,n} - p_{F_A,n}]}{\sigma_S^2 + |h_n|^2 p_{D,n} [1 - p_{D,n}]}, \tag{3.66} \]
which is clearly not the MRC result.

Note that the discussion above is only an attempt to refute, analytically, the idea that MRC is always optimal, regardless of the specific communication setting. We maintain that the above-obtained weights are for maximizing the MDC. In future extension of this chapter, we attempt to solve for the optimal weights and thresholds of problems (5.10) and (5.16).

In Fig. 3.9, we plot the resulting \( Q_m \) curve with respect to variations in the reporting SNR for the EGC, MRC, and the deflection coefficient based combining (DCC) schemes. For this curve, \( Q_f = 0.05 \), while the sensing SNR is fixed at 5dB. The figure demonstrates that the optimal weight vector for the problem of (5.10) remains different from the DCC and the EGC solutions. However, the suggested SEGC does perform better than the DCC for a large range of the reporting SNR.

\(^{11}\)We use lower-case boldface letters to denote vectors and upper-case boldface letters to denote matrices.
Figure 3.9: Variation of the minimum $Q_{\text{MD}}$ with the reporting SNR for the SEGC, MRC and the DCC weights using 20 sensors.

The MRC, on the other hand, is nowhere to be optimal in that figure. Because of the simplicity of its solution and ease of implementation, we chose to use the EGC method.
Chapter 4

An Optimal Probabilistic Multiple Access Scheme for Cognitive Radios

4.1 Background and Literature Review

Cross-layer design in cognitive radios has received much attention recently. For instance, the authors in [48] investigate the performance gains of cognitive radio networks under delay quality-of-service (QoS) limitations for the secondary users and an interference power constraint imposed by the primary user. Furthermore, the authors in [49] devise a cross-layer based opportunistic medium access (MAC) control protocol, which integrates spectrum sensing at the physical layer with packet scheduling at the link layer. Considering the problem of secondary power control, the authors of [50] characterize the tradeoff between the secondary transmission power, a physical layer concern, and the frequency of spectrum opportunities, a MAC layer issue, which impacts the secondary user scheduling policy.

Furthermore, [51] and [52] investigate an optimal scheduling policy for a secondary user coexisting with a primary user. The former transmits when the latter is inactive,
but uses the channel with some access probabilities if the primary is active. The design objective is to obtain the secondary user access probabilities to minimize the average queueing delay. Reference [53] employs the same framework as [51] providing a joint queueing analysis of the primary and the secondary queues. Both users are assumed to have equal priority, however. The assumed symmetry greatly simplifies the analysis, but is not suitable in the cognitive context.

In this chapter, we adopt a cross-layer design approach to the optimal scheduling of a cognitive user. We investigate the multiple access channel with one PU and one SU transmitting to a common receiver while using the same frequency channel. Previous work in the literature, for example [51], [52], [54], [55], and [56], also considered a PU and a SU communicating with a common receiver. In [54] and [55], a cognitive transmitter is allowed simultaneous access with the primary transmitter to a common destination using an appropriate space alignment design. The common receiver model is also used in [56] to solve a spectrum leasing problem consisting of maximizing the primary rate over a set of strategies that allow for secondary cooperation, but only under the constraint that minimum secondary rates are guaranteed. Herein, the common receiver may be the base station (BS) of a primary network that allows secondary users to utilize its resources. We assume the presence of an incentive that makes the BS permit secondary activity. This incentive can be extra revenue, especially when the primary network is not heavily loaded or when servicing an SU is critical. This revenue can come from the SUs and the spectrum regulator who is interested in maximizing the efficiency of spectrum usage. Another incentive can be cognitive relaying where the cognitive users aid the primary transmission by acting as relays [57], [58]. We investigate here the technical aspects of the postulated spectrum sharing model while other work in the literature, for example [59], focuses on the economical aspects.

Our goal is to devise an optimal scheduling policy for the SU in the presence of
sensing errors and block fading. Under a block fading channel model, channel inversion power control can be used to maintain communication at a fixed rate. Considering a similar spectrum sharing model, in [60] we implemented such power control for both users. However, under certain fading models, such as Rayleigh fading, channel inversion requires infinite power to maintain fixed rate communication [61]. Hence, herein we use truncated channel inversion to shut off transmission if the channel gain falls below a certain threshold. We then develop a queueing analysis for the SU taking into consideration the impact of imperfect sensing of the primary activity. Subsequently, we utilize results from our queueing analysis to optimize the SU scheduling policy so that the secondary queueing delay is minimized while constraining the primary outage probability below a specified maximum.

In our study, two schemes are presented for how the SU handles unsuccessful transmissions due to failure to detect primary activity and the subsequent collision with primary packets. In the first scheme, we assume that the secondary terminal does not attempt retransmission when collisions take place. The SU may be using a voice application that can tolerate some lost packets, but has strict delay constraints so that retransmissions are not done. The SU may also be employing across-packet coding so that data can be recovered even when some packets are lost [62]. The second scheme, on the other hand, is based on sustaining in the SU queue any packet that is received erroneously due to miss-detection of the PU. In this case, we assume that there is an acknowledgment feedback mechanism that informs the secondary transmitter of failed reception. The cognitive user clears the transmitted packet only if it receives feedback that the packet has been successfully decoded.

Initially, to facilitate understanding of the SU scheduling scheme, we consider no queue for the PU. We then investigate the scenario where the PU maintains a queue to buffer the packets that are lost due to collision with the SU’s transmission or that cannot be transmitted due to channel outage. The SU scheduling policy is then
modified to account for the primary queueing effects and to guarantee stability of the primary queue.

The contributions of this chapter can be summarized as follows: 

i) we provide a queueing analysis for a cognitive communication scheme that incorporates spectrum sensing errors and truncated channel inversion power control, 

ii) we solve a constrained queueing delay minimization problem to optimize the secondary access (the problem at hand is non-convex and no longer a simple linear program as in [60]),

iii) we introduce into our problem two schemes for handling secondary packet loss to cover a broader array of data applications, and

iv) we incorporate a primary queue to handle blocked or dropped transmissions, provide an approximate queueing analysis for the interaction between the two queues and also modify the SU queueing delay minimization problem to guarantee stability of the PU queue.

It is worthy to highlight that we endorse a cognitive model that allows cooperation only with the common receiver or BS of the primary network. We assume that no change is done to the way in which the PU operates. For example, the PU does not need to declare via a dedicated control packet at the beginning of the slot that it will transmit for the sake of aiding the SU to adjust its transmission parameters. It is the duty of the SU to implement spectrum sensing to assess primary activity. It is possible, of course, to make the PU help the SU more explicitly and, hence, obviate the need for sensing, but we think that it is more practical to assume that protocol modifications to accommodate the SUs are only confined to the common receiver and not all primary terminals.

The rest of the chapter is organized as follows. In Section 4.2 the system model is introduced with an investigation of the physical layer and the power control mechanisms employed by the PU and the SU. The impact of sensing errors is also presented. A queueing analysis is provided in Section 4.3 with a focus on obtaining the stationary distributions of the queues, which is in turn used to obtain expressions for the
average queueing delay and the average secondary transmit power. The constrained delay minimization problem is formulated in Section 4.4. In Section 4.5, we provide numerical results. Section 4.6 concludes the chapter.

4.2 System Model and Physical Layer

We consider a multiple access system comprised of a PU and an SU communicating with a common receiver. All terminals are equipped with a single antenna. As shown in Fig. 5.1, our setting can be viewed as a subsystem within a bigger network with different primary and secondary pairs using orthogonal frequency channels. We focus in this work on the pair using the same channel and, hence, can potentially interfere with one another.

System operation is slotted in time with the PU and the SU sending one packet per time slot when they transmit. Assuming that each SU packet has $L$ bits and that the time slot duration is $T$, we express the constant secondary transmission rate, $r_S$,
in units of bits per channel use, as

\[ r_S = \frac{L}{T W} = \log_2 (1 + K_S), \tag{4.1} \]

where \( W \) is the channel bandwidth and \( K_S \) is a constant signal-to-interference-and-noise ratio (SINR) term to be explained shortly. Since the PU may have a different packet size, and since it utilizes the whole slot without any sensing overhead as for the SU, the primary rate, \( r_P \), may be different from \( r_S \). Hence, we express the constant primary rate as

\[ r_P = \log_2 (1 + K_P), \tag{4.2} \]

where \( K_P \) is a constant SINR term to be explained shortly.

We allow the concurrent transmission of the PU and the SU packets through the use of superposition coding and successive interference cancellation (SIC) at the BS to subtract the SU’s interference from the PU’s received signal. This preserves the PU’s privileged access to the shared spectrum band with minimal interference and still provides a non-zero secondary rate. For the secondary signal to be decodable, the secondary transmission rate, \( r_S \), must be less than or equal to the capacity of the secondary link. When this condition is satisfied, the SU signal is decoded while treating the PU signal as interference. It is then subtracted before the PU signal is decoded.\footnote{For a survey of theoretical and practical aspects of superposition coding and single-antenna interference cancellation, please refer to [63,64].}

Note that due to the finite block size of the codewords, the decoding cannot be error-free. We consider this effect to be negligible relative to the outage caused when the secondary signal becomes undecodable due to sensing errors as explained below.

The SU senses the channel at the beginning of each time slot. The probability with which it transmits depends on the sensing outcome as explained in the following section. We take into account the impact of erroneous sensing, and consider the
events of false alarm, where the primary is sensed to be active while it is idle, and miss-detection, where the channel is sensed to be free despite primary activity. The introduction of sensing errors adds new challenges to the optimal user scheduling scheme at the BS.

A question arises into the model we consider here on the value of spectrum sensing by the SU albeit the consideration of a tightly cooperative cognitive system where there is a common receiver jointly decoding the PU and the SU messages. As mentioned in the Introduction, we postulate a cooperative system that requires no change in the method of operation of the PU. If the latter has a packet to send and its channel conditions are favorable, it will just transmit to the BS. The SU needs to ascertain primary activity to determine its transmission parameters at the beginning of the time slot. Rather than the PU having to announce that it is going to transmit, the SU senses the spectrum and makes a decision on primary usage of the communication channel. At the price of changing PU operation, the PU can aid the SU explicitly regarding its intent to transmit during a particular time slot. This would correspond to the perfect sensing scheme with no miss-detection or false alarm, which improves performance to the SU. In the numerical analysis section, we compare performance between the practical scheme with sensing uncertainties and the scheme with ideal sensing.

When the PU is OFF and the SU transmits, the signal at the BS receiver is \[ y = h_S x_S + n, \] (4.3)

where \( h_S \) is the SU-to-BS complex channel gain, assumed to be a random variable with some known distribution, \( x_S \) is the SU transmitted signal, and \( n \) is zero-mean additive white Gaussian noise (AWGN) assumed to be independent of the signal. We adopt a block fading model where the channel gain is assumed to be constant.
over the slot duration $T$, but changes independently from slot to slot. In the case of simultaneous primary and secondary transmissions, the received signal at the BS is given by

$$y = h_P x_P + h_S x_S + n,$$

(4.4)

where $h_P$ is the PU-to-BS complex channel gain, assumed to be a random variable with some known distribution, and $x_P$ is the PU transmitted signal. Define the primary and secondary channel gains by

$$\gamma_P = |h_P|^2, \quad \gamma_S = |h_S|^2,$$

(4.5)

with distributions $f_P(\gamma_P)$ and $f_S(\gamma_S)$, respectively, that depend on the statistics of $h_S$ and $h_P$. The two channel gains are assumed to be statistically independent. If frequency division duplexing (FDD) is considered, the BS estimates the channel gains from pilots embedded in the data packets transmitted by both the PU and the SU, and feeds back the channel estimates to the terminals. We assume that the estimates are perfect and are used by both the PU and the SU at the beginning of each time slot to determine their transmission power.

4.2.1 Primary and Secondary Power Control

In order to sustain a constant transmission rate, the PU and the SU are assumed to employ a truncated channel inversion mechanism, which is explained in [65] and investigated in a cognitive radio context in [66]. Using truncated channel inversion, the primary transmit power, $P_P$, is given by

$$P_P = \begin{cases} \frac{K_P \sigma^2}{\gamma_P} & \text{if } \gamma_P \geq \gamma^*_P, \\ 0 & \text{otherwise}, \end{cases}$$

(4.6)
where $\sigma^2$ is the noise variance and $\gamma_P^*$ is the channel gain threshold below which the primary transmitter remains silent. Since SIC is applied at the receiver to eliminate the secondary signal before decoding the primary message, the primary signal is decoded in the presence of AWGN only. Hence, the parameter $K_P$ of expression (4.2) is given by

$$K_P = \frac{P_P \gamma_P}{\sigma^2}. \quad (4.7)$$

As for the SU transmitter, it may transmit alone or concurrently with the primary. The secondary transmitter changes its power to make the rate fixed in both cases and for different secondary channel gains as long as they exceed a certain threshold $\gamma_S^*$. Let $P_F$ be the secondary transmit power when the primary is idle. Similar to the primary case, power $P_F$ is given by

$$P_F = \begin{cases} \frac{K_S \sigma^2}{\gamma_S} & \text{if } \gamma_S \geq \gamma_S^*, \\ 0 & \text{otherwise.} \end{cases} \quad (4.8)$$

On the other hand, the transmit power when the primary is also transmitting is expressed as

$$P_B = \begin{cases} \frac{K_S (P_P \gamma_P + \sigma^2)}{\gamma_S} & \text{if } \gamma_S \geq \gamma_S^*, \\ 0 & \text{otherwise.} \end{cases} \quad (4.9)$$

Note that the powers are functions of the channel gains though we do not use explicit notation for this dependence. Further, we note from the equations above that the application of a conventional channel inversion scheme implies that $\gamma_P^* = 0$ and $\gamma_S^* = 0$. This is not valid for certain channel fading models such as Rayleigh fading because, as we discuss later, the average secondary transmission power is proportional to $\mathbb{E} \left( \frac{1}{\gamma_S} \right)$, where $\mathbb{E}$ is the expectation operator. This quantity is infinite for a Rayleigh fading channel model, which makes it necessary to use a positive transmission power threshold according to the truncated channel inversion scheme.
We make the assumption that the SU knows the value of $\gamma_p^*$. If the primary channel is below cutoff, i.e., $\gamma_p < \gamma_p^*$, the primary remains silent and the SU then does not sense the channel. If the SU has a packet to transmit, it does so with probability one if $\gamma_s \geq \gamma_s^*$. We now study how the SU sensing errors affect packet transmission.

### 4.2.2 Impact of Imperfect Sensing

Under the false alarm event, the SU mistakenly detects the PU to be active with probability $p_{FA}$, hence it uses a higher power, $P_B$. Since the primary is OFF, the channel capacity of the secondary link when a false alarm occurs, expressed in bits/channel use, is given by

$$C_{S}^{FA} = \log_2 \left(1 + \frac{P_B \gamma_s}{\sigma^2}\right).$$  \hfill (4.10)

From (4.9), and assuming $\gamma_s \geq \gamma_s^*$, we have

$$\frac{P_B \gamma_s}{\sigma^2} = K_S \left(K_p \sigma^2 + \sigma^2\right) = K_S (1 + K_p).$$  \hfill (4.11)

Hence,

$$C_{S}^{FA} = \log_2 \left(1 + K_S [1 + K_p]\right) > \log_2 \left(1 + K_S\right) = r_S.$$  \hfill (4.12)

This means that the secondary signal is decodable at the BS. The problem here is that the SU uses more power than is required to ensure correct reception at the common receiver. In addition, since the channel is sensed busy, the SU may or may not access the channel depending on the values of the optimized access probabilities as explained in the next section. This means that false alarm may also cause the loss of a transmission opportunity, thus leading to an increase in delay.

The other sensing error event occurs when the SU mistakenly detects the PU to be

---

2 This knowledge can be made available through the common receiver.

3 Recall that the BS broadcasts the primary and secondary channel gains at the very beginning of the time slot.
inactive with probability \( p_{MD} \), thereby transmitting with power \( P_F \) to communicate its message. Consider the capacity of the secondary link when the primary activity is missed:

\[
C_{MD}^S = \log_2 \left( 1 + \frac{P_F \gamma_S}{P_P \gamma_P + \sigma^2} \right). \tag{4.13}
\]

It is straightforward to show that

\[
C_{MD}^S = \log_2 \left( 1 + \frac{K_S}{1 + K_P} \right) < \log_2 (1 + K_S) = r_S. \tag{4.14}
\]

Thus, because the secondary transmission rate exceeds the capacity of its link, the secondary signal becomes undecodable. We assume that this event causes the loss of both the primary and the secondary packets.

We summarize the cases of power transmission in Fig. 4.2. The figure presumes a certain channel realization for both \( \gamma_P \) and \( \gamma_S \) and, hence, some specific values for \( P_P, P_F \) and \( P_B \) as given in (4.6), (4.8) and (4.9). The first case corresponds to the PU transmitting alone with power \( P_P \), whereas in the second, the primary terminal is idle and is detected as such by the SU with probability \( 1 - p_{FA} \) or known to be idle because \( \gamma_P < \gamma_P^* \). The secondary transmission power is then \( P_F \). In the third case, the SU detects primary transmission with probability \( 1 - p_{MD} \) and transmits with power \( P_B \). The fourth and fifth cases depict the event of false alarm, where the SU transmits with \( P_B \) despite the idle state of the PU, and the event of miss-detection, where the PU is active but is not sensed as such by the SU which transmits with power \( P_F \). The figure excludes the case where both transmitters are silent.

### 4.3 The MAC layer

Now, we analyze the MAC layer setup for the primary and the secondary users. The SU has a finite buffer of length \( N \). We consider two cases for how the SU handles its
Figure 4.2: Cases of transmission in the cognitive multiple access system with sensing errors

packets. In the first case, we assume that any transmitted packet lost due to miss-detection is not considered for retransmission; the packet is simply discarded. This scenario practically models delay-sensitive data such as video, as considered in [62]. Hereafter, we refer to this operation mode as the delay-sensitive (DS) scheme. In the second scenario, we tackle the case where the SU’s packets, that are decoded in error by the secondary receiver due to miss-detection, are kept in the secondary queue for future retransmission. We refer to this scheme as the delay-tolerant (DT) scheme.

We distinguish between the primary and the secondary users by employing a priority structure such that the PU transmits with probability one when $\gamma_p \geq \gamma_p^*$ and a packet is available for transmission, while the SU transmits with probability one only if the PU is sensed idle and transmits with smaller probabilities otherwise. This is in addition to the physical layer constraint of decoding the PU’s signal after any possible interference from the SU is canceled. In this section, to highlight the SU queueing dynamics, we initially assume that the PU is working without a queue such that packets that cannot be transmitted due to channel outage or that are lost due to miss-detection are discarded. Afterwards, we consider the impact of the PU having a queue. Note that the PU without queue analysis mimics the delay sensitive nature of the SU but with a fundamental difference since the PU in this case simply does not have a queue.
4.3.1 SU Queueing Analysis

Let \( n_P(t) \) and \( n_S(t) \) denote the number of packets arriving to the PU and the SU at time slot \( t \), respectively. Assuming Bernoulli processes for the packet arrival patterns, we have \( n_P(t) = 1 \) with probability \( \alpha \), and \( n_S(t) = 1 \) with probability \( \theta \), where \( \alpha \) and \( \theta \) are the arrival probabilities of the PU and the SU, respectively. We adopt the early-arrival discrete-time queueing model where packets arrive at the beginning of the time slot \([67]\). Hence, if a packet arrives at a certain time slot and the queue is empty, the packet is transmitted during the same time slot. Hereafter, for compactness of presentation, we use the overbar notation to denote the probability of the complement of an event, for instance \( \overline{\alpha} = 1 - \alpha \) and \( \overline{\theta} = 1 - \theta \). The probability that the primary channel is sensed to be busy by the SU is

\[
\alpha^* = \alpha p_{MD} + \overline{\alpha} p_{FA}. \tag{4.15}
\]

Note that \( \alpha^* \) can be viewed as the PU arrival probability as perceived by the SU.

The SU maintains a buffer of finite length, \( N \), to store its packets. Let us denote the queue length at the end of slot \( t \) as \( q(t) \). Then, \( q(t) \) can be expressed as

\[
q(t) = \min \{ q(t-1) - u(t) + n_S(t), N \}, \tag{4.16}
\]

where \( u(t) \) is the number of packets transmitted at time slot \( t \). The SU’s transmission rate is fixed at one packet/time slot, thereby \( u(t) \in \{0,1\} \).

When the secondary channel gains exceeds the threshold, i.e. when \( \gamma_S \geq \gamma^*_S \), the SU decides to transmit in each slot with some probability based on the primary activity, the current queue length, and whether a new packet arrives. The SU transmits with probability one if the channel is sensed to be free. On the other hand, if the channel is found busy, the SU uses the following conditional access probabilities
The SU queue length, $q(t)$, can be described by an $N + 1$ birth-death Markov model\cite{68}, as shown in Fig. 4.3. The birth-death model is enforced by the fact that the SU does not receive nor transmit more than one packet in any given slot. In the Markov chain, the states represent the queue lengths. We now analyze this model.

**Proposition 1** The transition probabilities of the SU’s queue states are as follows:

(i) The probability of the queue length increasing by one, $\lambda_n = \Pr\{q(t) = n+1|q(t-1) = n\}$, is:

$$
\lambda_n = \begin{cases} 
\theta [\bar{\beta}_S + \beta_S \beta_P \alpha^* \bar{g}_n] & \text{for DS}^q, \\
\theta [\bar{\beta}_S + \beta_S \beta_P (\alpha^* \bar{g}_n + \alpha p_{MD})] & \text{for DT},
\end{cases}
$$

where $\alpha^*$ is given by (4.19), and where $\beta_P = \Pr\{\gamma_P \geq \gamma_P^*\}$ and $\beta_S = \Pr\{\gamma_S \geq \gamma_S^*\}$ denote the probabilities that the PU and the SU have favorable channel transmission conditions, respectively.

(ii) The probability of transitioning one state down, $\mu_n = \Pr\{q(t) = n-1|q(t-1) = \}

\footnote{Recall that DS and DT stand for the delay sensitive and delay tolerant schemes, respectively.}
\( n \), is:

\[
\mu_n = \begin{cases} 
\bar{\theta}\beta_S \left[ \bar{\beta}_P + \beta_P (\alpha^* f_n + \bar{\alpha}^*) \right] & \text{for DS}, \\
\bar{\theta}\beta_S \left[ \bar{\beta}_P + \beta_P (\alpha^* f_n + \bar{\alpha} \beta_{FA}) \right] & \text{for DT},
\end{cases}
\]  

(4.18)

where \( 1 \leq n \leq N \).

(iii) Finally, the probability of self-transition, \( \Pr \{ q(t) = n | q(t-1) = n \} \), for either the DS or the DT case is:

\[
1 - \lambda_0 \text{ for } n = 0, \quad 1 - \lambda_n - \mu_n \text{ for } 1 \leq n \leq N - 1,
\]

and \( 1 - \mu_N \text{ for } n = N \).

\[ (4.19) \]

**Proof.** The proof for (i) follows from the fact that the queue length is incremented by one in the DS case if there is a new secondary packet and if there is no secondary transmission. The SU does not transmit when \( \gamma_S < \gamma_S^* \), which happens with probability \( \beta_S \). If \( \gamma_S \geq \gamma_S^* \), there is no secondary transmission with probability \( \bar{\beta}_n \) when \( \gamma_P \geq \gamma_P^* \) and the primary is perceived as active. In the DT case, in addition to the previous DS conditions, we add the condition that secondary transmission has not been successful due to collision with primary transmission, hence the term \( \alpha p_{MD} \).

The proof for (ii) is based on the fact that since an unsuccessfully received packet is discarded under the DS mode of operation, the queue length is decremented by one if there is no new secondary packet arrival and the SU transmits, regardless of the transmission outcome. In the DT case, the queue length is decremented by one if the SU transmits successfully and there is no new secondary packet arrival. In the DS case, when the primary is perceived to be idle with probability \( \bar{\alpha}^* = 1 - \alpha^* \), the SU transmits with probability one. On the other hand, the DT expression has the term \( \bar{\alpha} \beta_{FA} \) because successful transmission dictates that the primary is idle and is correctly perceived as such. A miss-detection causes a decoding failure and the transmitted packet is kept in the secondary queue under the DT scheme.
Finally, the proof of (iii) simply follows from the requirement that the transition probabilities from a state sum up to unity.

Given the previous proposition, we state the following results.

- **Stationary Distribution:** The stationary distribution, \( \pi_i \)'s, of the \( N + 1 \) Markov chain for the SU’s queue can be obtained as follows:

\[
\pi_n = \pi_o \prod_{k=0}^{n-1} \frac{\lambda_k}{\mu_{k+1}}, \text{ where: } \pi_o = \left( 1 + \sum_{n=1}^{N-1} \prod_{k=0}^{n-1} \frac{\lambda_k}{\mu_{k+1}} \right)^{-1}. \tag{4.20}
\]

Note that the use of the stationary queue distribution is justified by the BASTA (Bernoulli arrivals see time averages) property \[69,70\]. Equation (4.20) is equivalent to the local balance equations of the Markov chain:

\[
\pi_n \lambda_n = \pi_{n+1} \mu_{n+1}, 0 \leq n \leq N - 1. \tag{4.21}
\]

- **Outage Probability:** The probability of primary outage, which is the same for the DS and the DT schemes, is given by

\[
p_{\text{out}} = \alpha \left[ \beta_\text{p} + \beta_\text{p} \beta_\text{s} p_{\text{MD}} (\theta \pi_o + \pi_o) \right]. \tag{4.22}
\]

Outage is defined as the event of having a packet to send but failing to transmit it successfully. Note that \( \alpha \beta_\text{p} \) is the primary outage in the absence of secondary operation. The term including \( p_{\text{MD}} \) represents the impact of miss-detection, with \( (\theta \pi_o + \pi_o) \) corresponding to the probability that the SU has a packet to transmit.

- **Buffer Loss Probability:** The probability of blocking a packet from entering the secondary queue represents the buffer loss probability. The latter is defined in the DS case as the probability of the buffer being full and there is a new arrival while
the SU decides not to transmit. This probability is given by

\[ p_{DS}^L = \pi_N \theta \left[ \beta_S + \beta_S \beta_P \alpha^* g_N \right]. \]  \hspace{1cm} (4.23)

Under the DT scheme, the buffer loss probability is the probability of the buffer being full while there is a new arrival and the SU does not transmit or transmits unsuccessfully. This loss probability is given by

\[ p_{DT}^L = \pi_N \theta \left[ \beta_S + \beta_S \beta_P (\alpha^* g_N + \alpha p_{MD}) \right]. \]  \hspace{1cm} (4.24)

The additional term for the DT scheme comes from the design requirement that in the event of miss-detection, the SU retains the packet in its queue for future retransmission.

- **Packet Delay:** Using Little’s law [68], the average packet delay, expressed in terms of the SU arrival rate, the loss probability \( p_L \), which can be \( p_{DT}^L \) or \( p_{DS}^L \), and the stationary queue state distribution, is given by:

\[ D = \frac{1}{\theta p_L} \sum_{n=0}^{\infty} n \pi_n. \]  \hspace{1cm} (4.25)

In light of the loss probabilities given by (4.23) and (4.24), we set \( g_N = 1 \) to decrease the loss probability. This has the impact of making \( p_{DS}^L = \pi_N \theta \beta_S \) and \( p_{DT}^L = \pi_N \theta [\beta_S + \beta_S \beta_P \alpha p_{MD}] \).

- **Packet Loss Rate** The overall packet loss rate, \( \zeta_L \), is the probability of losing a packet. For the DT scheme:

\[ \zeta_{DT}^L = p_{DT}^L, \]  \hspace{1cm} (4.26)

when a packet is transmitted unsuccessfully under the DT scheme, it is kept in the queue. A packet is only lost due to buffer overflow. Under the DS scheme, in addition to the \( p_{DT}^L \) term, packets are lost during the course of transmission due to collision
with missed primary packets. The probability of this event is
\( \beta_S \beta_P \alpha \eta p_{MD} (\theta \pi_o + \pi) \)
and, hence,
\[
\zeta_{DS}^L = p_{DS}^L + \beta_S \beta_P \alpha \eta p_{MD} (\theta \pi_o + \pi).
\] (4.27)

### 4.3.2 SU Power Consumption

Using the Markov chain modeling presented in the previous subsection, we propose
the following for the average power consumption of the SU.

**Proposition 2** Let \( \mathbb{E}(P_F) \) be the average of \( P_F \) over the channel gains, and let \( \tilde{P}_S^{DS} \)
and \( \tilde{P}_S^{DT} \) denote the average SU transmission power normalized by \( \mathbb{E}(P_F) \) under the
DS and the DT schemes, respectively. These power levels are given by

\[
\tilde{P}_S^{DS} = \bar{\phi} + (1 + K_P) \left( \frac{\theta}{\beta_S} - \bar{\phi} \right) + \pi_o \beta K_P \bar{\phi} - \pi_N \theta (1 + K_P) \frac{\beta_S}{\beta_S},
\] (4.28)

\[
\tilde{P}_S^{DT} = \bar{\phi} + (1 + K_P) \left( \frac{\theta}{\beta_S} - \eta \right) + \pi_o \beta ((1 + K_P) \eta - \bar{\phi})
- \pi_N \theta (1 + K_P) \left( \frac{\beta_S}{\beta_S} + \alpha p_{MD} \beta_P \right),
\] (4.29)

where \( \bar{\phi} = 1 - \beta_P \alpha^* \) and \( \eta = \beta_P + \beta_P \alpha^* p_{FA} \).

**Proof.** In the Appendix 4.7., we detail the proof for the DT scheme. The proof
for the DS scheme can be done in a similar fashion by using the appropriate transition
probabilities.

Thus far, we have discussed the SU queue assuming that the PU operates without
a queue. We now investigate the SU scheduling problem assuming that the PU
maintains a queue to keep packets that arrive when the primary channel is in outage
or that need to be retransmitted because of collision induced by miss-detection.
4.3.3 Optimal Scheduling with the PU queue

If the PU and the SU both have queues, the proper analysis is to use a 2-dimensional Markov chain in order to model both queues and their interaction. In this chapter, we develop an approximate analysis that decouples the primary and secondary queueing analyses and allows for a tractable optimization problem. In section 4.5, we validate the analytical results presented here via an actual system simulation.

We assume that the PU has an infinite buffer. The queue transition probabilities are given as follows. The probability that the primary queue increases by one is

\[
\lambda_P = \alpha \left[ \beta_{sp} + \beta_{spMD} (\theta \pi_o + \pi_0) \right],
\]

which is the same as (4.22). The primary queue length is incremented by one if there is a new primary packet, which happens with probability \(\alpha\), and the primary cannot transmit because \(\gamma_P < \gamma^*_P\) or cannot transmit successfully due to miss-detection by the SU. If the queue is nonempty, the probability of the queue length decreasing by one, denoted \(\mu_P\), is given by

\[
\mu_P = \bar{\alpha} \left( 1 - \left[ \beta_{sp} + \beta_{spMD} (\theta \pi_o + \pi_0) \right] \right),
\]

because a decrement by one requires that no new primary packet arrives and that there is a successful transmission. For a queue length \(k > 0\), if \(\epsilon_k\) is the probability of the primary queue being in state \(k\), we have the following balance equation: \(\epsilon_k (\mu_P + \lambda_P) = \epsilon_{k-1} \lambda_P + \epsilon_{k+1} \mu_P\). Since \(\epsilon_0 \lambda_P = \epsilon_1 \mu_P\), it can be shown that \(\epsilon_k = \epsilon_o \left( \frac{\lambda_P}{\mu_P} \right)^k\). Furthermore, given that \(\sum_{k=0}^{\infty} \epsilon_k = 1\), we have

\[
\epsilon_o = 1 - \frac{\lambda_P}{\mu_P}.
\]

Note that for the sum \(\sum_{k=0}^{\infty} \epsilon_k\) to exist, the term \(\frac{\lambda_P}{\mu_P}\) must be less than unity. This
means that for a stationary distribution to exist for the Markov chain, the condition 
\( \lambda_p < \mu_p \) must hold. This is the condition for the primary queue stability [71]. This 
is equivalent to having \( \alpha + \beta_p + \beta_p \beta_{SPMD} (\theta \pi_o + \pi_o) < 1 \), which can be written as

\[
\alpha + \frac{\lambda_p}{\alpha} < 1 \rightarrow \lambda_p < \alpha \bar{\alpha}.
\] (4.33)

The probability of primary transmission is given by

\[
\alpha_{\text{eff}} = \alpha \epsilon_o + 1 - \epsilon_o.
\] (4.34)

Indeed, the PU transmits if its queue is non-empty or if the queue is empty but a
new packet arrives at the very beginning of the time slot. In the next section, we
formulate a linear programming optimization problem in order to obtain the optimal
SU’s access probabilities. We also explain how (4.34) can be incorporated in the
optimization problem when a primary queue is considered.

### 4.4 Queueing Delay Minimization

Our objective is to minimize the SU delay subject to a certain PU outage probability
and an average SU power constraint. The optimization variables are \( \gamma^*_S \), and the
transmission probabilities \( f_n \) and \( g_n \) for queue state \( n \). As for the primary channel
threshold gain, \( \gamma^*_P \), it is obtained as follows. Based on (4.6), the average primary
power, \( \mathbb{E}(P_P) \), can be expressed as:

\[
\mathbb{E}(P_P) = \mathbb{E}_{\gamma^*_P} \left( \frac{1}{\gamma^*_P} \right) K_P \sigma^2,
\] (4.35)

\footnote{Similar to the SU, we adopt here the early-arrival queueing model [67].}
where $\mathbb{E}_{\gamma_p} \left( \frac{1}{\gamma_p} \right) = \int_{\gamma_p}^{\infty} \frac{1}{\gamma_p^2} f_p(\gamma_p) d\gamma_p$ and $\overline{\gamma}_p$ is the average primary channel gain. We assume that $\mathbb{E}(P_p)/\sigma^2$ and $r_p$ are specified. Parameter $K_p$ is then obtained from (4.12). Equation (4.35) can then be numerically solved to obtain $\gamma_p^*$. 

### 4.4.1 Delay Minimization without Primary Queue

Our queueing delay minimization problem can be formulated for the DS and the DT cases as:

$$
\text{minimize } D = \frac{1}{\theta_p L} \sum_{n=1}^{N} n \pi_n, \tag{4.36}
$$

where $\pi_n$ is given by (4.20), and where $p_L$ is given by $p_L^{DS} = \pi_N \theta_S$ and $p_L^{DT} = \pi_N \theta (\beta_S + \beta_p \alpha p_{MD})$ in the DS and the DT cases, respectively, when $g_N = 1$. The delay minimization objective is subject to the following constraints:

$$
\begin{align*}
    p_{out} &\leq p_{out}^{(max)}, \tag{4.37} \\
p_{L} &\leq p_{L}^{(max)}, \tag{4.38} \\
\hat{P} &\leq P_S, \tag{4.39} \\
\lambda_n^- &\leq \lambda_n \leq \lambda_n^+, \tag{4.40} \\
\mu_n^- &\leq \mu_n \leq \mu_n^+, \tag{4.41} \\
\pi_0 = \left(1 + \sum_{n=1}^{N} \prod_{k=0}^{n-1} \frac{\lambda_k}{\mu_{k+1}} \right)^{-1}, \tag{4.42} \\
\pi_n = \pi_0 \prod_{k=0}^{n-1} \frac{\lambda_k}{\mu_{k+1}}, \tag{4.43}
\end{align*}
$$

where $p_{out}$ is given by (4.22), $p_{out}^{(max)}$ as mentioned in the previous section is the maximum allowable primary outage probability, $p_{L}^{(max)}$ is the maximum allowable buffer loss probability, $\hat{P}$ is given by $\hat{P}_S^{DS}$ in (4.28) or $\hat{P}_S^{DT}$ in (4.29) for the DS and the
DT cases, respectively, and $P_S$ is a specified limit on the average SU power. Since
the access probabilities $0 \leq f_n \leq 1$ and $0 \leq g_n \leq 1$, the transition probabilities are
bounded, as represented by the constraints (4.40) and (4.41). The upperbounds and
the lowerbounds on the transition probabilities can be readily obtained from (4.17)
and (4.18) as:

$$
\lambda_n^+ = \begin{cases} 
\theta \left[ \beta_S + \beta_S \beta_P \alpha^* \right] & \text{for DS}, \\
\theta \left[ \beta_S + \beta_S \beta_P \left( \alpha^* + \alpha_{p_{MD}} \right) \right] & \text{for DT}, 
\end{cases}
$$

(4.44)

$$
\lambda_n^- = \begin{cases} 
\theta \beta_S & \text{for DS}, \\
\theta \left[ \beta_S + \beta_S \beta_P \alpha_{p_{MD}} \right] & \text{for DT},
\end{cases}
$$

$$
\mu_n^+ = \begin{cases} 
\theta \beta_S & \text{for DS}, \\
\theta \beta_S \left[ \beta_P + \beta_P \left( \alpha^* + \alpha_{p_{FA}} \right) \right] & \text{for DT},
\end{cases}
$$

(4.45)

$$
\mu_n^- = \begin{cases} 
\theta \beta_S \left[ \beta_P + \beta_P \alpha^* \right] & \text{for DS}, \\
\theta \beta_S \left[ \beta_P + \beta_P \alpha_{p_{FA}} \right] & \text{for DT}.
\end{cases}
$$

Before explaining how the optimization problem is solved, we investigate the de-
pendency of the queueing delay on the zero-state probability, $\pi_o$. We first obtain
the range of the feasible $\pi_o$ from the constraints of the delay minimization problem.
Using (4.22), the PU outage probability constraint given by (4.37) can be expressed
as a constraint on the minimum possible value of $\pi_o$. Specifically,

$$
\pi_o \geq \pi_{o,\min} = \max \left\{ 0, \frac{\beta_P \beta_S \alpha_{p_{MD}} + \alpha \beta_P - \beta^\text{(max)}_{out}}{\alpha_{p_{MD}} \theta \beta_P \beta_S} \right\}.
$$

(4.46)

On the other hand, the SU average power expressions, (4.28) or (4.29), together with
the SU power constraint shown in (4.39) impose a maximum possible value on $\pi_o$ such that $\pi_o \leq \pi_{o,\text{max}}$, where $\pi_{o,\text{max}}$ is given by

$$
\min \left\{ \frac{P_S - \phi + (1 + K_P) \left[ \pi_N \theta \left( \frac{\eta}{\beta_s} \right) + \phi - \frac{\eta}{\beta_s} \right]}{\theta K_p \phi} \right\} \quad (4.47)
$$

in the DS case, and by

$$
\min \left\{ \frac{P_S - \bar{\phi} + (1 + K_P) \left[ \pi_N \theta \left( \frac{\eta}{\beta_s} + \alpha \rho_{\text{MD}} \beta_p \right) + \eta - \frac{\eta}{\beta_s} \right]}{\theta ((1 + K_P) \eta - \bar{\phi})} \right\} \quad (4.48)
$$

for the DT case, wherein $\eta = \beta_p + \beta_p \alpha \bar{p}_{\text{FA}}$. Thus, the probability $\pi_o \in [\pi_{o,\text{min}}, \pi_{o,\text{max}}]$.

According to (4.36), the delay is minimized by minimizing $\pi_1, \pi_2, \ldots, \pi_N$. The probability normalization constraint dictates that $\sum_{n=1}^{N} \pi_n = 1 - \pi_o$. As such, increasing $\pi_o$ decreases the sum of the non-negative probabilities $\pi_n$ for $n \geq 1$. Since the delay is a weighted sum of $\pi_n$'s with positive weights, then the minimum delay is achieved by maximizing $\pi_o$. Note that maximizing $\pi_o$ reduces $\pi_N$ which decreases the buffer loss probability, $p_L$, further reducing delay. Henceforth, we use $\pi_o = \pi_{o,\text{max}}$.

In order to solve the optimization problem above efficiently, we convert it to the following equivalent form. First, define

$$
\rho_n = \prod_{k=0}^{n-1} \lambda_k \mu_{k+1}, \quad 1 \leq n \leq N. \quad (4.49)
$$

Note from the above equation that $\rho_n$ is non-negative for all $n$. Now (4.43) can be written as

$$
\pi_n = \pi_o \rho_n. \quad (4.50)
$$

Then, summing from $n = 1$ to $n = N$ and using total probability normalization to unity, we attain the following equality:

$$
\sum_{n=1}^{N} \rho_n = \pi_o^{-1} - 1. \quad (4.51)
$$
Further, we use the upperbound and the lowerbound on $\lambda_n$ and $\mu_n$, respectively, in (4.44) and (4.45), to write

$$\rho_1 = \frac{\lambda_0}{\mu_1} \leq \rho_+,$$

where $\rho_+$ is given by:

$$\rho_+ = \begin{cases} \frac{\theta [\beta_S + \beta_S \beta_P \alpha^*]}{\beta_S [\beta_S + \beta_P \pi]} & \text{for DS}, \\ \frac{\theta [\beta_S + \beta_S \beta_P \alpha_{MD}]}{\beta_S [\beta_S + \beta_P (\alpha^* + \pi \rho F_A)]} & \text{for DT}. \end{cases}$$

Then, making use of (4.21), we obtain

$$\rho_{n+1} = \frac{\lambda_n}{\mu_{n+1}} \rho_n \leq \rho_+ \rho_n, \quad 0 \leq n \leq N - 1.$$  \hspace{1cm} (4.54)

We also use the lowerbound and the upperbound on $\lambda_n$ and $\mu_n$, respectively, given in (4.44) and (4.45), to write

$$\rho_1 = \frac{\lambda_0}{\mu_1} \geq \rho_-, \hspace{1cm} (4.55)$$

where $\rho_-$ is given by:

$$\rho_- = \begin{cases} \frac{\theta \pi_S}{\beta_S} & \text{for DS}, \\ \frac{\theta [\beta_S + \beta_S \beta_P \alpha_{MD}]}{\beta_S [\beta_S + \beta_P (\alpha^* + \pi \rho F_A)]} & \text{for DT}. \end{cases}$$

Now using (4.50), the delay can be written as:

$$D = \frac{1}{\theta \rho_L} \sum_{n=1}^{N} \pi_o n \rho_n = \frac{1}{\theta \rho_L} \pi_o \sum_{n=1}^{N} n \rho_n.$$  \hspace{1cm} (4.57)

Note in the above equation that $p_L = \pi_N \theta \beta_S$ in the DS case and $p_L = \pi_N [\beta_S + \beta_S \beta_P \alpha_{MD}]$ for the DT case, i.e., delay depends on $\pi_N$. Also, $\pi_o$ depends on $\pi_N$ from (4.47) and (4.48) and, subsequently, the $\rho_n$'s, $1 \leq n \leq N$, depend on $\pi_N$ from (4.50).
Algorithm 2 Delay minimization algorithm

1: Construct sets $A$ and $B$ for the attempted $\pi_N$ and $\gamma^*_S$ values, respectively.
2: for $\pi_N \in A$ do
3:  for $\gamma^*_S \in B$ do
4:    Solve the linear program given by (4.58) under constraints (4.59).
5:  end for
6: end for
7: Obtain the minimum delay and the optimal $\pi_N$, $\gamma_S$ and $\{\rho_n\}_{n=1}^N$.
8: Obtain the optimal transmission probabilities $f_n$ and $g_n$ as explained in Section 4.5.

This makes our optimization problem non-convex. Nevertheless, observe that if $\gamma^*_S$ and $\pi_N$ are fixed, the optimization problem becomes linear in both the objective and the constraints, thereby becoming a linear program that can be efficiently solved.

Hence, the problem can be solved by doing a two-dimensional grid search over $\gamma^*_S$ and the feasible region of $\pi_N$ and solving a linear program for each pair to determine the minimum delay. This solution is facilitated by the fact that for a given $\gamma^*_S$ and $\pi_N$, the value of $\pi_\circ$ is fixed at $\pi_{0,\text{max}}$. The linear program given $\pi_N$ and $\gamma^*_S$ has the following form:

$$
\text{minimize } D = \frac{\pi_\circ}{\theta p_L} \sum_{n=1}^N n \rho_n \\
\text{subject to:}
$$

$$
\sum_{n=1}^N \rho_n = \pi_\circ^{-1} - 1, \quad \rho_1 \leq \rho_+, \quad \rho_1 \geq \rho_-,
$$

$$
\rho_{n+1} - \rho_+ \rho_n \leq 0, \quad \rho_{n+1} - \rho_- \rho_n \geq 0, \quad \rho_n \geq 0,
$$

where $\rho_+$ and $\rho_-$ are given by (4.53) and (4.56), respectively. Algorithm 2 summarizes how the non-convex delay minimization problem is solved.

Extensive simulations have shown that $\pi_N$ is practically zero when the SU buffer size, $N$, is large enough. We can use this to find an approximate solution for the delay minimization problem, with reduced complexity. We can set $\pi_N = 0$ and do
a one-dimensional search over $\gamma^*_S$ solving the linear program for each value of $\gamma^*_S$. The approximate solution for large buffer sizes lends itself to the following analytical solution.

Note from (4.36) that the queueing delay is the summation of $\rho_n$’s, each weighted by $n$. This means that the minimum delay can be achieved by maximizing the $\rho$’s with small $n$, subject to the non-negativity constraint and the equality constraint

$$\sum_{n=1}^{N} \rho_n = \frac{1}{\pi_o} - 1.$$  

Since $\rho_1 \geq \rho_-$ and $\rho_{n+1} > \rho_n \rho_-$, then in order to have a feasible solution we should have $\rho_- + \rho_2^2 + \cdots \rho_N^N \leq \frac{1}{\pi_o} - 1$ otherwise, the equality constraint will dictate that some $\rho$ values go below the minimum. If the inequality $\rho_- + \rho_2^2 + \cdots + \rho_N^N < \frac{1}{\pi_o} - 1$ is satisfied, we find the optimal solution starting from $\rho_1, \rho_2, \text{till } \rho_{N-1}$ as follows.

Now, assume that we want to determine the value of $\rho_n$. We want it to be the maximum possible such that it still satisfies $\rho_n \leq \rho_{n-1} \rho_+$ for $n > 1$ and $\rho_1 \leq \rho_+$, and also the condition that

$$\rho_n + \rho_n \rho_- + \rho_n \rho_-^2 + \cdots + \rho_n \rho_-^{N-n} \leq \frac{1}{\pi_o} - 1 - \sum_{m=1}^{n-1} \rho_m,$$  

such that the minimum value constraints on $\rho_{n+1}$ to $\rho_{N-1}$ are not violated. This means that for $n > 1$, we have

$$\rho_n = \min \left\{ \rho_{n-1} \rho_+, \frac{1}{\pi_o} - 1 - \sum_{m=1}^{n-1} \rho_m \right\},$$  

and

$$\rho_1 = \min \left\{ \rho_+, \frac{1}{\pi_o} - 1 \right\}.$$  

(4.61)
4.4.2 Delay Minimization with Primary Queue

Now we explain how the presence of a primary queue alters the delay minimization problem. Probability $\alpha_{\text{eff}}$ in (4.34) can be written as:

$$\alpha_{\text{eff}} = \frac{\alpha}{\beta_p \left[ 1 - \beta_s \pi_{MD} (\theta \pi_o + \pi_o) \right]}, \quad (4.63)$$

which can be expressed in a more compact form as $\alpha_{\text{eff}} = \frac{\alpha}{1 - \frac{\lambda_p}{\alpha}}$. Note that, per (4.33), stability requires the condition $\lambda_p < \alpha \pi$. This makes $\alpha_{\text{eff}} < 1$. A conservative approach for design is to use the maximum allowable PU system outage probability, $P_{\text{out}}^{(\text{max})}$, to replace $\lambda_p$. This makes $\alpha_{\text{eff}}$ independent of $\pi_o$ and facilitates the analysis. Accordingly, $\alpha_{\text{eff}}$ is given by

$$\alpha_{\text{eff}} = \frac{\alpha}{1 - \frac{P_{\text{out}}^{(\text{max})}}{\alpha}}, \quad (4.64)$$

This value of $\alpha_{\text{eff}}$ should be used in lieu of $\alpha$ due to the incorporation of the infinite primary queue analysis in the previous SU MAC analysis in subsection [4.3.1]. That is, we solve the same optimization problem as in the latter subsection but with each $\alpha$ replaced by $\alpha_{\text{eff}}$ given by (4.64). This design is conservative because it uses the maximum specified primary outage probability $P_{\text{out}}^{(\text{max})}$ whereas the actual outage may be less than this value. The SU becomes more conservative in its transmission policy because it assumes it inflicts the maximum possible outage on the PU.

Alternatively, we may compute $\alpha_{\text{eff}}$ iteratively by considering both the SU and the PU queues as detailed in Algorithm 3. As shown in the next section, the value of $\alpha_{\text{eff}}$ obtained iteratively is always less than or equal to the value obtained from (4.64). This improves the secondary throughput at the expense of increased computational complexity to determine the optimal access parameters due to iterations. However, it is important to mention that our simulations show that the value of $\alpha_{\text{eff}}$ converges after a few iterations only.
Algorithm 3 Delay minimization algorithm with a primary queue

1: Initialize $\alpha_{\text{eff}} = \alpha$.
2: repeat
3:  Solve Algorithm 2 with $\alpha_{\text{eff}}$ to obtain $\pi_0$.
4:  Use $\pi_0$ and (4.32) to obtain $\epsilon_0$.
5:  Use $\epsilon_0$ and (4.34) to update $\alpha_{\text{eff}}$
6: until change in the value of $\alpha_{\text{eff}}$ is below a specified tolerance.
7: Check the stability of the primary queue.

4.5 Numerical Results

In this section, we solve the linear program under the DS and the DT schemes and provide numerical results and comparisons in terms of average queueing delay and packet loss rate. The channels between the users and the BS are Rayleigh fading channels. Hence, the channel gains are exponentially distributed: $f_P(\gamma_P) = \frac{1}{\gamma_P} \exp\left(-\frac{\gamma_P}{\gamma_P}\right)$ and $f_S(\gamma_S) = \frac{1}{\gamma_S} \exp\left(-\frac{\gamma_S}{\gamma_S}\right)$. We use $\overline{\gamma}_P = 1$ and $\overline{\gamma}_S = 1$. When $E(P_P)/\sigma^2 = 10$ and $K_P = 10$, (4.35) can be used to obtain $\gamma^*_P = 0.265$ and, consequently, $\beta_P = 0.767$. The SU buffer size is set to $N = 50$, unless otherwise stated, the PU arrival rate $\alpha = 0.4$, the system loss probability $p_{l,\text{max}} = 0.025$, and the PU outage probability $p_{\text{out},\text{max}} = 0.025 + \alpha \beta_P$. Note that the maximum outage the SU is allowed to make in this case is $p_{\text{out},\text{max}} - \alpha \beta_P = 0.025$. The secondary parameters $K_S = 10$ and $P_S = 1/\int_{\gamma_S}^{\infty} \frac{1}{\gamma_S} f_S(\gamma_S) d\gamma_S$. The sensing error probabilities are $p_{\text{MD}} = 0.1$ and $p_{\text{FA}} = 0.1$, unless otherwise specified.

Fig. 4.4 shows the difference between the average queueing delay versus $\theta$ for the DS and the DT schemes. As expected and because of its operating mechanism, the DS scheme experiences a smaller expected delay relative to its DT counterpart for a given secondary arrival rate, $\theta$. Also plotted on this graph is the approximate solution for the DT case with $\pi_N$ set to zero, implemented via the analytical solution of (4.61) and (4.62). It is clear that for the relatively large buffer size, $N = 50$, the difference between the exact and the approximate solutions is negligible, albeit the computational cost difference is significant. The improved delay performance
Figure 4.4: Comparison of the performance of the DS and DT schemes versus the secondary arrival rate $\theta$. Also plotted is the approximate solution with $\pi_N = 0$.

achieved by the DS scheme comes at the expense of an increased packet loss rate compared to the DT scheme as evident from Fig. 4.5. The overall packet loss rate for the DT case, $\zeta_{L}^{\text{DT}}$, is almost zero, whereas $\zeta_{L}^{\text{DS}}$ increases with $\theta$ approaching the bound $p_{\text{out}}^{(\text{max})} - \alpha \beta_P$. Recall that under the DS scheme, when the PU loses a packet due to a miss-detection, the SU also loses its packet as it is discarded from the queue to enhance the queueing delay performance.

Fig. 4.6 illustrates the dependence of the queueing delay on the miss-detection and false alarm probabilities. Increasing the sensing error probabilities causes an increase in the queueing delay. As described earlier, we also compare the case of no sensing errors to the various values of miss-detection and false alarm probabilities. For the given system parameters, it is noted from the figure that when sensing is reliable, e.g. $p_{\text{FA}} = p_{\text{MD}} = 0.01$, the performance of the SU is not different from what it could achieve if the PU where to cooperate, and this would require change in the PU protocol. Fig. 4.7 illustrates the variation of the PU outage probability with the SU arrival rate for different values of sensing errors. It is noted that the prescribed outage probability of 2.5% is never exceeded. Similar to Fig. 4.6, the case of ideal sensing is also considered for comparison. Fig. 4.8 demonstrates the impact of the
Figure 4.5: Comparison of the overall packet loss rate for the DT and the DS schemes.

Figure 4.6: Variation of queueing delay with the sensing error probabilities in the DT scheme.
buffer size $N$ on the queueing delay. For small $\theta$, the delay is almost the same for $N = 5$, 10 and 50. As $N$ decreases, the delay rises sharply at larger values of $\theta$.

In Fig. 4.9, we show the optimal scheduling parameters for $N = 10$. Note that the $\rho_n$’s we obtain from the solution of the optimization problem are unique, however they are ratios of the optimal $\lambda_n$’s and $\mu_n$’s that map to the scheduling probabilities $f_n$ and $g_n$. Hence, $f_n$ and $g_n$ values may not be unique. We set these access probabilities to 0 or 1 whenever possible in order to facilitate implementation. For the figure parameters, it is feasible to set all $g_n$’s to unity and compute the corresponding $f_n$’s.
Figure 4.9: Values of the optimal scheduling parameters, $f_n$, for the case of $g_n = 1$ ($\theta = 0.5$, $\alpha = 0.4$, $N = 10$, and $p_{MD} = p_{FA} = 0.1$).

Figure 4.10: Queueing delay as a function of the arrival rate $\theta$ for $\alpha = 0.3$ assuming no primary queue. This is contrasted with the delay for the case of a primary queue using Algorithm 3 or conservative design based on (4.64). Also validated are the analytical formulas of Algorithm 3 with system simulation results.
Figure 4.11: Fraction of 5000 simulations with varied system parameters for which the relative error in $\pi_0$ is less than or equal to a certain value. The error refers to the discrepancy between simulation results and those obtained using Algorithm (3).

Fig. 4.10 shows the queueing delay as a function of the arrival rate $\theta$ when we assume that the PU has an infinite queue and operates in a DT-like mode. The figure provides the delay when no queue is assumed for the PU and $\alpha = 0.3$. This is contrasted with the delay for the case of a primary queue, with an original $\alpha = 0.3$ and using the conservative design of (4.64). It is evident from the figure that when we account for the PU queueing effects, the delay is increased for the SU since the effective primary rate increases. Also shown in the figure is the delay when $\alpha_{\text{eff}}$ is computed using Algorithm 3. The delay in this case is lower than that corresponding to the conservative design. Because we resorted to an approximate analysis of the PU and the SU queueing interaction, we also validate the iterative solution used to obtain $\alpha_{\text{eff}}$ by a system simulation. As observed, the simulation results are in close agreement with our proposed solution.

Finally in Fig. 4.11 we plot the relative error between the analytical results for $\pi_0$ obtained iteratively using Algorithm 3 and the simulated results. We verify that the percentage difference between our proposed approach and the actual system simulation is no more than 1% in 90% of the simulations. Note that we obtain this
figure using 5000 simulations with varied system parameters. Similar results exist for the $\epsilon_o$ values.

4.6 Conclusions

Considering the practical situation of imperfect spectrum sensing and time-varying fading channels, we developed in this chapter an optimal access policy for a SU sharing the same spectrum band with a PU and communicating with a common receiver. The optimality criteria here refers to minimizing the SU queueing delay under average transmit power and PU outage probability constraints. Regarding how the collided primary and secondary packets are handled in the system, we devised two strategies: discard collided packets because of the delay-sensitive nature of the data, or buffer the packets until future retransmission opportunities. As expected, we showed that the delay-sensitive case achieves a better delay performance than the delay-tolerant one, albeit at the expense of increased packet loss.

We also considered the more realistic scenario of the PU maintaining a queue to the case where no queue was assumed. We showed that this increases the apparent primary packet arrival rate from the point of view of the SU. This leads to an increased delay, which sometimes can be significant as illustrated numerically. This shows the importance of including the primary queueing effects on cognitive operation.

4.7 Appendix

[Proof for $P_{DS}^S$, the Average SU transmission power]Proof for $P_{DS}^S$, the Average SU transmission power normalized by $\mathbb{E}(P_F)$

We begin by noting that using (4.17) and (4.18) we can obtain the SU’s access
probabilities $f_n$ and $g_n$ as follows:

\[
\begin{align*}
    g_n &= 1 + \frac{\alpha P_{MD}}{\alpha^*} - \frac{1}{\alpha^* \beta S \beta P} \left[ \frac{\lambda_n}{\theta} - \beta_S \right], \\
    f_n &= \frac{1}{\alpha^* \beta P} \left[ \frac{\mu_n}{\theta} - \beta_P \bar{\alpha} - \beta_P \bar{\alpha} P_{FA} \right].
\end{align*}
\]  

(4.65)

(4.66)

Now, we write an expression for the transmit power when the queue length is $n$, denoted $P_{DT}(n)$, averaged over the primary and secondary channel gains. For $n > 0$, we have

\[
P_{DT}(n) = \int_{\gamma_S}^{\infty} f_S(\gamma_S) \frac{K_S \sigma^2}{\gamma_S} \left[ \beta_P + \beta_P \bar{\alpha}^* \right] \\
+ \psi_P \left( \theta g_n + \bar{\theta} f_n \right) d\gamma_S
\]

(4.67)

\[
= K_S \sigma^2 \left[ \int_{\gamma_S}^{\infty} f_S(\gamma_S) d\gamma_S \right] \\
\times \left[ \bar{\phi} + \psi_P \left( \theta g_n + \bar{\theta} f_n \right) \right].
\]

For $n = 0$, we have

\[
P_{DT}(0) = \theta K_S \sigma^2 \left[ \int_{\gamma_S}^{\infty} f_S(\gamma_S) d\gamma_S \right] \times \left[ \bar{\phi} + \psi_P g_0 \right],
\]  

(4.68)

where $\psi_P = \beta_P \alpha^* (1 + K_P)$. An explanation of the terms in (4.67) and (4.68) is in order. The primary channel is in outage with probability $\bar{\beta}_P$, and the primary is perceived by the SU to be inactive with probability $\beta_P \bar{\alpha}^*$. Under these two cases, the SU transmits—when it has a packet to send—with power $\frac{K_S \sigma^2}{\gamma_S}$ so that it achieves the secondary transmission rate $r_s = \log_2 (1 + K_S)$. If the gain of the primary channel exceeds $\gamma_P^*$ and the primary is sensed active by the SU, the latter transmits with a probability that depends on both the queue state and whether there is a new
secondary packet arrival. The transmission probability in this case is given by \( \theta g_n + \bar{\theta} f_n \). Regardless of the actual activity state of the PU, the SU transmits with a power 
\[
K_S \frac{P_p \gamma + \sigma^2}{\gamma_S} = K_S \frac{\kappa + \sigma^2}{\gamma_S}
\] 
to sustain the transmission rate \( r_S \) under the belief that the PU is active. When the SU queue is empty, there is no secondary transmission except when there is a new packet arrival. Hence, in (4.68), there is only \( g_0 \) and \( P_{DT}(0) \) is proportional to \( \theta \).

Thus, the normalized secondary average transmit power, \( \hat{P}_S^{DT} \), is given by

\[
\hat{P}_S^{DT} = \frac{\sum_{n=0}^{N} P^{DT}(n) \pi_n}{\mathbb{E}(P_F)} = \frac{\sum_{n=0}^{N} P^{DT}(n) \pi_n}{K_S \sigma^2 \int_{\gamma_S}^{\infty} f(\gamma_S) \, d\gamma_S}.
\] (4.69)

When we set \( g_N = 1 \) to minimize the buffer loss probability, we obtain

\[
\hat{P}_S^{DT} = \pi_o \left[ \bar{\phi} + \psi P g_0 \right] + \pi_N \left[ \bar{\phi} + \beta P \alpha^* \left( \theta + \bar{\theta} f_N \right) \right] + \sum_{n=1}^{N-1} \pi_n \left[ \bar{\phi} + \psi (\theta g_n + \bar{\theta} f_n) \right].
\] (4.70)

Since \( \sum_{n=1}^{N-1} \pi_n \bar{\phi} = \bar{\phi} (1 - \pi_o - \pi_N) \), and using equation (4.65), \( \hat{P}_S^{DT} \) can further be expressed as

\[
\hat{P}_S^{DT} = \bar{\phi} - \pi_o \lambda_0 \frac{1 + K_P}{\beta_S} + \pi_N \mu_N \frac{1}{\beta_S} + \pi_o \left( \theta \psi P \left[ 1 + \frac{\alpha P M D}{\alpha^*} + \frac{\beta_S}{\alpha^* \beta_S} \right] - \bar{\phi} \right) + \pi_N \psi P \left( \theta - \frac{\bar{\theta}}{\alpha^* \beta_P} \left[ \beta_P + \beta_P \alpha P M D \right] \right) + \theta \psi P \sum_{n=1}^{N-1} \pi_n \left[ 1 + \frac{\alpha P M D}{\alpha^*} - \frac{1}{\alpha^* \beta_S} \left( \frac{\lambda_n}{\theta} - \beta_S \right) \right] + \bar{\theta} \psi P \sum_{n=1}^{N-1} \pi_n \frac{1}{\alpha^* \beta_P} \left[ \mu_n \beta_P - \beta_P \alpha P M D \right].
\] (4.71)

The summation in (4.71) has the term \( \frac{1 + K_P}{\beta_S} \left( \sum_{k=1}^{N-1} \pi_k \mu_k - \sum_{k=1}^{N-1} \pi_k \lambda_k \right) \). Combining
this term with $-\pi_0 \lambda_0 \frac{1+K \beta}{\beta S} + \pi_N \mu_N \frac{1+K \beta}{\beta S}$, we obtain the term $\sum_{k=1}^{N} \pi_k \mu_k - \sum_{k=0}^{N-1} \pi_k \lambda_k$, which is equal to zero because of the queue balance equation shown in (4.21). The remaining terms can be manipulated in a straightforward manner to yield (4.29). Proceeding in a similar manner, equation (4.28) can also be verified.
Chapter 5

Enhanced Primary and Secondary Performance Through Cognitive Relaying and Leveraging Primary Feedback

5.1 Background and Literature Review

In a typical CR scenario, the cognitive node makes an access decision on the basis of the spectrum sensing outcome. The problem with this approach is that sensing the primary transmitter activity does not inform the secondary terminal about the impact it has on the primary receiver (PR), which is the terminal that needs to be protected from the secondary’s interference. This issue has induced interest in employing the feedback from the primary receiver to its respective transmitter in designing cognitive access protocols [72–75].

In [72], for example, the ARQ feedback is exploited to design a retransmission-based secondary access policy. The policy is randomized and aims at maximizing the SU throughput, with a constraint on the throughput loss and failure probability of the
PU. As opposed to what we do in this work, no sensing is involved at the secondary side, notwithstanding the authors claim that a channel sensing based approach would be suboptimal to their retransmission based policy. In fact, although [72] highlights the importance of exploiting the primary ARQ feedback, it is clearly suboptimal to rely only on partial observations such as the ARQ since the spectrum sensing observations can also be used for a more refined interference management in CR networks. Furthermore, although sensing is employed in [73], it is considered to be perfect, which is an unrealistic assumption given the challenging aspects of spectrum sensing [7]. In this context, an investigation of a cognitive access policy which is based on both channel sensing and retransmission feedback while taking into account practical issues is a must, a goal that is central to this chapter.

On the other hand, while many works in the open literature advocate strict transparency of the primary network to cognitive operation, such transparency cannot be maintained given the realities of spectrum sensing and given the potential benefit that the primary network can reap from the key concept of cooperation. Cognitive relaying is one form of such cooperation that has been considered in many papers, for instance [76–81]. The main idea of a cooperative cognitive user appeared in [76, 77], where the maximum stable SU throughput is investigated while guaranteeing stability of the primary queue. Cooperation consists of the SU acting as a relay for the undelivered packets of the PU. In order to aid the primary transmission, the SU has an additional relaying queue to store primary data packets. An acceptance probability is used to control the number of primary packets that are allowed into the relaying queue at the secondary node.

Also, two secondary and one primary nodes are investigated in [78]. All nodes communicate with the same receiver employing a random access scheme. Opportunistic sensing is employed where the SUs sense the channel at different times depending on the quality of the channel to the common receiver. Retransmissions are allowed,
but the feedback information is not incorporated in the analysis. The design criterion is to construct bounds on the stability region of the average secondary arrival rates so that all queues in the system, primary and secondary, are stable. In [79], multiple SUs are considered together with one PU. The SUs form a cluster with a common relaying queue to forward the undelivered packets of the PU. In [79], the maximum stable throughput for the SUs is sought given a specified primary throughput. However, perfect sensing is assumed and although retransmissions are allowed, the feedback information is also not incorporated in the analysis. On the other hand, in [80] the authors characterize the stable-throughput region in a two-user cognitive shared channel with multi-packet reception (MPR) capability added to the physical layer of the nodes and cognitive relaying. The SU in [80] bases its transmission actions on the status of the primary queue. It is not clear, however, how such information can be acquired.

In this chapter, we present a randomized secondary access policy that is based on both spectrum sensing and the primary ARQ feedback. The use of feedback information is exploited here to refine the imperfect sensing results. Furthermore, and unlike the model we presented in [75], the secondary transmitter (ST) in this work acts as a relay for the PU in the event of transmission failure on the direct link. This cooperative protocol is implemented to achieve performance enhancements for both the primary and secondary users despite the sensing errors. The ST maintains a queue to store primary packets in order to relay them when necessary. By studying the behavior of the original queue at the primary transmitter (PT) and the one at the ST which stores the correctly received primary packets, we formulate expressions for the primary and the secondary throughputs and for the primary queueing delay. We obtain the optimal secondary access probabilities via maximizing the SU’s throughput while guaranteeing the stability of the considered queues, a minimum primary throughput and a maximum primary packet delay. Furthermore, while in [75] the
sensing time was negligible, herein we incorporate the sensing time and investigate its impact on the sensing errors and the achievable secondary throughput.

While previous works, e.g. \[72,73,82\], take a partially observable Markov decision process (POMDP) approach to derive optimal policies to their formulated problems, we propose a randomized secondary access policy, under the assumption of stationarity and ergodicity. In our approach, the SU maps the outcome of the sensing and feedback processes to fixed probabilities that take into account time average performances of the involved queueing systems. Our approach is similar to the one proposed in \[83\], albeit in a different context. This choice of policy is motivated by a desire to simplify the proposed algorithm versus a computationally intense POMDP formulation which not only maps randomized policies on the whole state space, i.e. ARQ and queueing state, but also requires knowledge of the state of the system, e.g. knowledge of state transition probabilities and cost functions, or at least the ability to learn such state. Furthermore, we quantify the loss of optimality in our approach by comparing it with an upperbound system where the state of the PU is perfectly known, such as in the case of a perfect sensing scenario with negligible sensing time.

In conclusion, the contributions of our work may be summarized as follows. We design an optimal secondary access scheme that is based on both the primary ARQ feedback and the spectrum sensing outcome. We expand the notion of protection for the PUs by incorporating multiple QoS performance metrics such as the throughput and the queueing delay in addition to the stability of the primary queues. We also study the impact of the sensing time on the sensing errors and demonstrate a sensing throughput tradeoff. Our results indicate the benefits associated with cognitive relaying for both the primary and the secondary terminals.

In detailing these contributions, the rest of this chapter is organized as follows. An overview of the proposed system and the physical layer model are presented in Section \[5.2\]. In Section \[5.2.2\] we introduce the probabilistic secondary access scheme.
and in Section 5.3, we analyze the Markov chains associated with the primary queues in the network. In Section 7.4, we solve the secondary throughput maximization problem subject to primary throughput and stability constraints. We analyze the delay associated with the cognitive relaying scheme and incorporate it as an additional constraint in the optimization problem in Section 5.5. Section 6.5 discusses illustrative numerical results, followed by the conclusion in Section 5.7.

5.2 System Model

We consider a time slotted system consisting of one PU and one SU communicating with their respective receivers over the same frequency channel as shown in Fig. 5.1. As argued in previous work in the literature, e.g. [76][80][84], our setting can be viewed as a subsystem within a bigger network with different primary and secondary pairs using orthogonal frequency channels. We focus in this work on the basic building block of the network.

![Figure 5.1: Cognitive multiple access scheme employing feedback: A PU and a SU share the same frequency resource.](image)

When a terminal transmits during a time slot, it sends exactly one packet to its respective receiver. The primary terminal implements an ARQ error control protocol. Error checking at the receiver uses the cyclic redundancy code (CRC) bits attached to each packet. Transmission success or failure is declared by sending back to the
transmitter either an acknowledgment (ACK) or negative acknowledgment (NACK) on a feedback channel. The ARQ protocol is untruncated, i.e., there is no maximum on the number of retransmissions and an erroneously received packet is retransmitted until it is received correctly. Moreover, we assume that the primary ARQ feedback is un-encrypted and is available to the ST. The feedback is assumed to be received correctly by the PT and the ST due to the use of a strong channel code.

The SU solves for the optimal medium access probabilities so as to maximize its throughput provided that a minimum PU throughput and a maximum queueing delay are guaranteed. The access probabilities depend on the spectrum sensing outcome and the latest ARQ feedback emitted from the PR and overheard by the ST. If the ST senses the channel, it consumes \( \tau_s \) seconds in the process, thereby reducing the amount of time available for packet transmission. However, it still transmits one packet per time slot by adjusting its transmission rate. This can be implemented, for example, by using a signal constellation with more symbols or by increasing the channel coding rate or both. Note that by doing this, the probability of link outage increases as explained in Appendix 5.8.1. If the ST does not transmit, it receives the primary data packet transmitted by the PT when its queue is nonempty.

The PU maintains an infinite buffer to store its data packets. We call this buffer the PU queue. The ST has a designated queue that can be used to store the primary packets and forward them to the PR. This is dubbed the PS queue. The ST accepts a primary packet with a certain probability in case it receives the packet correctly and the PR does not. The acceptance probability provides a degree of freedom that can be used to optimize the system and ensure queue stability. In addition, at any given time slot, the ST makes a choice between transmitting its own packets, with the optimized access probabilities, or relaying the PU’s stored packets.

We describe below the physical layer model first and then the MAC layer setup, leading to the formulas we use to solve the SU throughput maximization problem.
5.2.1 Physical Layer

We assume independent block fading for the channel gains between any pair of nodes. Outage occurs when the transmission rate exceeds the channel capacity. We distinguish between two outage events depending on whether the packet is decoded in noise only or in noise-plus-interference. Further, the ST may transmit at the beginning of the time slot or after a delay $\tau_s$ if it performs spectrum sensing for a duration $\tau_s$ in order to ascertain the state of primary activity.\footnote{Whether the ST chooses to sense or not is part of the probabilistic spectrum access policy and will be shortly explained in the next subsection.}

We introduce the following outage probabilities to characterize our proposed system. Probabilities $\delta_S(0)$ and $\delta_S(\tau_s)$ denote the outage probability of the secondary link when the PT is silent in the two cases of direct transmission without sensing and transmission after $\tau_s$ seconds, respectively. If the PT transmits concurrently with the ST, the secondary outage probabilities are $\Delta_S(0)$ and $\Delta_S(\tau_s)$. It should be the case that $\Delta_S(0) \geq \delta_S(0)$ and $\Delta_S(\tau_s) \geq \delta_S(\tau_s)$ due to the interference from the PT. Similarly, the outage probability of the primary link is $\delta_P$ when the ST is idle, and $\Delta_P$ when the ST transmits simultaneously with PT.

Due to the cooperation between the primary and secondary users, there are also two more relevant links: the link between the PT and the ST over which the ST can receive the primary packets, and the link between ST and PR over which the ST can relay the primary packets to the PR. We assume that the ST is a half-duplex node, meaning that if the SU transmits, it cannot simultaneously receive a packet and perform spectrum sensing. Therefore, we need one only outage probability for the PT-ST link, which we denote as $\delta_{PS}$, since there can be no concurrent transmission due to the half-duplex assumption. Likewise, the outage probability of the ST-PR link is $\delta_{SP}(0)$ when the ST transmits at the beginning of the time slot and $\delta_{SP}(\tau_s)$ if it undergoes spectrum sensing first, as explained in detail in the next subsection.
assume that if there are concurrent transmissions from the PT and the ST to the PR, the latter can only decode the packet it gets from PT. The aforementioned outage probabilities are a function of the number of bits in a data packet, the slot duration, the channel bandwidth, the transmit powers and the average channel gains, as detailed in Appendix 5.8.1. These parameters are assumed known by ST.

Errors in spectrum sensing at the ST are characterized by two probabilities: $p_{FA}$, which corresponds to the false alarm case where an idle PU is erroneously detected to be active, and $p_{MD}$, which pertains to the missed detection case where the ST observes the channel to be free while the PU is actively transmitting. Probabilities $p_{FA}$ and $p_{MD}$ are a function of the sensing time $\tau_s$ (cf. Appendix 5.8.2. for expressions of the sensing error probabilities when an energy detector is used). In the numerical analysis section, we optimize over $\tau_s$ and demonstrate the existence of a sensing-throughput tradeoff.

5.2.2 MAC Layer

At each time slot, a packet arrives at the PU queue with probability $\theta$, where $\theta$ is the packet arrival probability. The primary arrival process is independent from one time slot to the other. The ST is back-logged, i.e. it always has packets to send. The ST overhears the primary ARQ feedback represented by a short packet that is sent at the end of the time slot reporting the decoding status at the PR. We consider two ARQ feedback schemes:

- **Explicit ACK Scheme**: In this case, the PR sends an ACK whenever the primary packet is decoded properly. Hence, the absence of an ACK due to transmission failure or the inactivity of the PT is interpreted as a NACK by the SU.

---

2This assumption is for analytical tractability of our problem since otherwise the operation of the PU and the PS queues cannot be decoupled. The performance of our system serves as a lower bound of any other counterpart where the PR decodes the signal from both the PT and the ST.
• *Explicit NACK Scheme:* Here, whenever there is a transmission error, a NACK is fed back to the PT. Absence of feedback due to correct reception or inactivity is interpreted by the ST as an ACK.

The use of the explicit ACK or explicit NACK schemes is determined by the primary network. Secondary operation follows how the original primary ARQ feedback is implemented. If the last received primary ARQ feedback is a NACK under the explicit NACK approach or there is no primary feedback in the explicit ACK approach, the ST accesses the channel at the beginning of the next time slot with probability $p_N$. In this case, the SU does not sense the channel and assumes that the primary terminal will be active during the next time slot retransmitting its data packet. If, on the other hand, the last primary feedback is interpreted as an ACK, then the primary may or may not transmit in the next time slot depending on the availability of a packet to send. In this case, the ST senses the channel first and accesses the channel with probability $p_F$ if the channel is sensed to be free, and with probability $p_B$ if the channel is sensed to be busy.

Here, we devise a probabilistic access scheme with the deterministic scheme of having the access probabilities, $p_N$, $p_F$ and $p_B$, equal to 0 or 1 as a special case. Regardless of the feedback scenario, the three access probabilities, $p_N$, $p_F$ and $p_B$, are obtained via maximizing the secondary throughput given a constraint on the minimum acceptable primary throughput and maximum tolerable primary queueing delay.

If the last received primary ARQ is interpreted as an ACK by the SU, i.e. the secondary will sense the channel in the next time slot, the probability of secondary transmission *while the primary is active*, $p_A$, is given by

$$p_A = p_{MD} p_F + p_{MD} p_B,$$  \hspace{1cm} (5.1)
where $\pi = 1 - x$. Note that there is no difference between the explicit ACK and NACK schemes when the PU is active.

In addition to incorporating the ARQ feedback of the primary and the sensing results into its access probabilities, the SU acts a relay for the PU in the event of decoding failure on the direct primary link. The SU accepts failed primary packets with probability $\nu$ in case it receives such packets correctly. This probability can be tuned in order to control the level of cooperation between the PU and the SU and, at the same time, it ensures the stability of the primary queue at the SU. Note that setting $\nu = 0$ corresponds to the case of no primary-secondary cooperation which we covered in [75].

The SU maintains a queue for the PU packets it receives, the PS queue, where the failed packets are stored to be forwarded to the primary receiver when the primary user is sensed to be silent. Because the SU now acts as a relay for the primary packets, the ARQ protocol operates as follows.

- *Explicit NACK*: If a NACK is fed back by the PR but the primary packet is decoded successfully and accepted at the ST, then the ST overrides the NACK with an ACK to signal to the PT that it can eliminate the packet from the PU queue and that the ST is responsible for delivering the packet to the PR. The accepted packet is stored in the ST’s PS queue.

- *Explicit ACK*: If no ACK is heard from the PR but the primary packet is decoded successfully by the ST and admitted to the PS queue, then the ST transmits an ACK to signal to the PT that it can remove the packet from the PU queue.

In summary, a transmitted primary packet is kept in the PU queue for retransmission when the PR does not decode it and the ST either decodes it erroneously or

---

3Recall that if the ST does not transmit, it operates in the receiving mode during the time slot so that it can receive the primary transmission in case the PU queue is nonempty.
does not accept it. The packet is removed from the PU queue when either the PR
decodes it successfully or the ST accepts it into the PS queue after correct decoding.
The length of the PS queue at the ST is incremented by one when the PR fails to
decode a packet while it is successfully decoded and accepted by the ST. It is im-
portant to note that the secondary network is not fully transparent to the primary
network, since some protocol changes are needed to account for the modified feedback
because of the secondary relaying. For example, if a NACK is fed back by the PR
but the primary packet is decoded successfully and accepted at the ST, then the ST
may use a mini-slot on a dedicated primary control channel to send this feedback
to the PT [86]. Such a measure is not expected to have a significant time cost on
the transmission time because of the short duration of the mini-slot, it does however
require cooperation between the primary and the secondary networks. Nevertheless,
as mentioned in the Introduction, this comes at the advantage of enhanced perfor-
mance for the primary network such as a higher throughput, which we demonstrate
numerically in Section 6.5.

The PU queue follows the early-arrival model, which means that when a packet
arrives, it does so at the very beginning of the time slot [67]. So, if the buffer is
empty and if a packet arrives, it can be transmitted in the same time slot during
which it has been received. The PS queue, on the other hand, operates according to
the late-arrival model since in this discrete-time queueing model the packet arrives
late into the time slot. If the buffer is empty and a new packet is received, it is
transmitted during the next time slot [67]. The PS queue conforms to the late-arrival
model because primary packets would not be accepted into the PS queue until they
have been received in error by the PR first. Since this event occurs near the end
of the time slot, as per the error control protocol, there is not enough time for the
packet to be transmitted during the remainder of the time slot.
Let $\overline{\beta}$ be the probability of correct primary packet reception or packet removal from the PU queue when the last primary feedback is interpreted as an ACK at the secondary side. We then have the following expression for $\overline{\beta}$:

$$
\overline{\beta} = p_A \overline{\Delta P} + \overline{p}_A \overline{\delta P} + \overline{p}_A \delta_P \overline{\delta PS} \nu.
$$

(5.2)

The explanation of (5.2) is that following an ACK, the secondary terminal may transmit in the next slot with probability $p_A$, given by (5.1), causing the probability of primary correct transmission to be $\overline{\Delta P}$. If the SU does not access the channel, which occurs with probability $\overline{p}_A$, the probability of correct reception by the primary receiver is $\overline{\delta P}$. Finally, the term $\overline{p}_A \delta_P \overline{\delta PS} \nu$ corresponds to the probability that the ST successfully accepts a primary packet that the PR has failed to receive. Recall that we assume that the SU is a half-duplex node, meaning that if the SU transmits, it cannot simultaneously perform spectrum sensing and receive a packet. Proceeding in a similar vein, we have the following expression for $\overline{\gamma}$, the probability of correct primary packet reception or packet removal from the PU queue when the last primary feedback is interpreted as a NACK by the ST:

$$
\overline{\gamma} = p_N \overline{\Delta P} + \overline{p}_N \overline{\delta P} + \overline{p}_N \delta_P \overline{\delta PS} \nu.
$$

(5.3)

Next, we consider the technical details of the PU and the PS queues where the formulas of (5.2) and (5.3) will be used to analyze the transition probabilities for the Markov chains we describe next.
5.3 Queueing Analysis

5.3.1 PU Queue: The Primary Queue at the PT

We model the primary queue at the PT in the presence of the SU by the Markov chain illustrated in Fig. 5.2. Figures 5.3 and 5.4 show the specific initial states under the explicit ACK and the explicit NACK schemes, respectively. Note that we have two classes of states. The first class is the ‘A’ states which are reached after an ACK, whereas the second class is the ‘N’ states which are preceded by a NACK. It is important to note that the ACK or NACK we refer to here may be explicit or implicit depending on the primary system operation. The main difference though between the explicit ACK and the explicit NACK schemes lies in the initial states depicted in Figs. 5.3 and 5.4.

Figure 5.2: Markov chain representing the primary queue given the last received primary ARQ. ‘A’ states are the ones entered into after an ACK is received, whereas ‘N’ states are visited after a NACK. When the queue has a non-zero number of packets, there is no difference between the explicit ACK and explicit NACK cases.

Figure 5.3: Same as Fig. 5.2 but with a focus on the states with \( k = 0, 1 \) and 2 for the explicit ACK case. Note that the transition between state \((0, A)\) to \((0, N)\) has a probability \(\theta\) because if no packet arrives, the primary remains idle. Since, in the explicit ACK approach, the lack of an ACK is interpreted as a NACK, the Markov state makes a transition to state \((0, N)\).

Let the probability of the ‘A’ states when the queue has \( k \) packets be \(\pi_k\), while the
Figure 5.4: Same as Fig. 5.3 but for the explicit NACK scheme, where a NACK is sent when the packet is decoded unsuccessfully at the primary receiver. The absence of NACK means either that there has been no transmission or that the sent packet has been received correctly. Hence, state \((0, A)\) makes a transition to itself when there is no packet to transmit and \(\eta_0\), the probability of being in state \((0, N)\), is zero.

The probability of the ‘N’ states is \(\eta_k\). Self transitions are not shown in the figures. Note that a state \((k, A)\) with \(k \geq 1\) makes a self transition with probability \(\theta_\beta\) because the primary successfully delivers a packet, but has its queue at the same state due to the arrival of a new packet. The state \((k, N)\) with \(k \geq 1\) moves to itself with probability \(\bar{\theta}_\gamma\) as the packet is transmitted in error and there is no new primary packet. Similar reasoning can be applied to understand the remaining depicted transition probabilities. We have shown in \([75]\) that \(\pi_0, \eta_0, \pi_k, \eta_k, \sum_{k=1}^{\infty} \pi_k\) and \(\sum_{k=1}^{\infty} \eta_k\) can be written as summarized in Table 5.1 for the explicit ACK and explicit NACK cases.

We recall that a queue is said to be stable if and only if the probability of being empty remains nonzero for a time \(t\) that grows to infinity \([71]\). Hence, the condition for the stability of the PU queue is that \(\pi_0 > 0\).

5.3.2 PS Queue: The Buffer at the ST for Storing Primary Packets

Let \(a_{PS}\) be the packet arrival probability at the PS queue. Then, we have

\[
a_{PS} = \delta_P \delta_{PS} \nu \left[ \bar{p}_A \left( \pi_0 \theta + \sum_{k=1}^{\infty} \pi_k \right) + \bar{p}_N \left( \eta_0 \theta + \sum_{k=1}^{\infty} \eta_k \right) \right]. \tag{5.4}
\]
Table 5.1: Steady state probabilities for the PU queue Markov chain under the explicit ACK and explicit NACK schemes. Parameter $\Psi = \theta^2 + \bar{\theta}^2$.

The explanation of the above expression is as follows. For an arrival event to occur at the PS queue, the PU should transmit a packet while the SU should remain silent, all while the primary packets fail to be decoded properly at their intended receiver. The PU transmits a packet when either the PU queue contains a packet or, if its buffer is empty, when a new primary packet arrives. The SU remains silent with probability $p_A$ following an ACK, and with probability $p_N$ following a NACK. Finally, the primary packet fails to be decoded properly at the PR, whereas it is received correctly and accepted by the ST with probability $\delta_P \delta_{PS} \nu$.

We assume that the ST, if it decides to transmit during a time slot, makes a choice between transmitting a packet from the PS queue, if it is non-empty, or transmitting its own packets. We use $p_{PS}$ to denote the probability of transmitting from the PS queue. The success transmission probability, $b_{PS}$, from the PS queue is thus,

$$b_{PS} = p_{PS} \left[ \pi_0 \bar{\theta} (p_F p_F + p_{FA} p_{PB}) \delta_{SP}^{(\pi_0)} + \eta_0 \bar{\theta} p_N \delta_{SP}^{(0)} \right],$$

where $\delta_{SP}^{(0)}$ is the probability of correct reception by the PR of a packet transmitted by the ST, given that the PT is OFF and the ST starts transmission immediately at the beginning of the time slot. If the ST spends $\tau_s$ seconds to sense the channel and
the PT is OFF, the probability of correct reception by the PR is $\delta_{SP}^{(\tau_s)}$. As previously mentioned, we assume here that if the PR now receives two packets simultaneously, one from the PT and one from the ST, it correctly decodes the packet from the PT only with probability $\Delta_P$. Hence, the successful secondary forwarding to the PR requires that the PT remains silent. This means that the PU buffer should be empty and there is no new packet arrival. This is the term between square brackets in (5.5).

Note that if not transmitting from the PS queue, the ST transmits its own packets with probability $\bar{p}_{PS}$.

The discrete-time PS queue may be represented by a Birth-Death Markov chain. Let $\psi_k$ denote the probability that the PS queue contains $k$ packets. By a simple analysis of the local balance equations, we can write the following expressions for the PS queue state probabilities:

$$\psi_k = \frac{\psi_0 \left[ a_{PS} \bar{b}_{PS} \right]^k}{\bar{b}_{PS} \left[ a_{PS} \bar{b}_{PS} \right]^k}, \quad (5.6)$$

where,

$$\psi_0 = 1 - \frac{a_{PS}}{\bar{b}_{PS}}. \quad (5.7)$$

Similar to the PU queue, the condition for the stability of the PS queue is that $\psi_0 > 0$.

Here, it is important to mention that the evolution of the PU and the PS queues is decoupled. While the parameters of the PS queue depend on the PU queue, the reverse is not true. This is primarily due to our assumption that if the PR now receives two packets simultaneously, one from PT and the other from ST, it correctly decodes the packet from the PT only with probability $\Delta_P$. Further, we have verified, via system simulation, the validity of this decoupling as highlighted in the numerical results section.
5.4 Optimization Framework

We now characterize the secondary and primary throughputs, which we use to obtain the optimal SU access probabilities, $p_F$, $p_B$ and $p_S$ and also to obtain the optimal $p_{PS}$ and $\nu$ values. For compactness of presentation, we provide formulas for the throughput that are generic for the explicit ACK and explicit NACK cases, and show in the numerical analysis section the difference between both cases under the relaying capability.

Here, we define throughput as the probability of successful transmission provided that the terminal has a data packet to send. Recall that the SU is backlogged, that is, it always has a packet to send. Based on our analysis above, the secondary throughput formula may be written as:

$$R_s = \left( \psi_0 \bar{p}_{PS} + \psi_0 \right) \left( \sum_{k=1}^{\infty} \pi_k + \theta \pi_0 \right) p_A \Delta_S^{(\tau_s)} + \eta_0 \bar{\theta} p_N \delta_S^{(0)}$$

$$+ \left[ \sum_{k=1}^{\infty} \eta_k + \theta \eta_0 \right] p_N \Delta_S^{(0)} + \pi_0 \bar{\theta} (\bar{p}_{FA} p_F + p_{FA} p_B) \delta_S^{(\tau_s)}.$$  \hspace{1cm} (5.8)

Note that the term $\psi_0 \bar{p}_{PS} + \psi_0$ represents the probability that the ST sends a secondary packet. This occurs if the PS queue is empty, or with probability $\bar{p}_{PS}$ if the PS queue is nonempty, which happens with probability $\psi_0$. The remaining terms on the right-hand-side of the equation indicate the possibilities of secondary transmission while the PT is active or idle. The summations in (5.8) are provided in Table 5.1 for the explicit ACK and explicit NACK cases.

Given the definition of throughput as the probability of successful transmission \textit{provided} that the terminal has a data packet to send, it is straightforward to show that the primary throughput is represented by:

$$R_p = \frac{\theta}{1 - \bar{\theta}(\pi_0 + \eta_0)},$$  \hspace{1cm} (5.9)
where the denominator indicates the probability that the PT has a packet to send.

Now in order to obtain the optimal access probabilities, we solve the following optimization problem:

$$\max_{p_F, p_B, p_N, p_{PS}, \nu} R_s$$

subject to:

$$R_p \geq R_p^*, \quad \pi_0 > 0, \quad \psi_0 > 0,$$

(5.10)

where $R_p^*$ is the minimum guaranteed primary throughput. We also constrain the optimization problem to guarantee stability of the PU and the PS queues. Note that, using (5.9), the minimum rate requirement constraint can be re-written as $\pi_0 > \frac{1}{\theta} \left(1 - \frac{\theta}{R_p^*}\right)$ for the explicit NACK case and as $\pi_0 > \frac{\theta}{\theta} \left(1 - \frac{\theta}{R_p^*}\right)$ for the explicit ACK case. Hence, when $R_p^* > \theta$, the rate requirement constraint effectively guarantees the stability of the PU queue.

It can be verified that the optimization problem in (5.10) is non-convex. Hence, the approach to exactly solve the above problem is to do a search over the optimization variables’ values, which are all probabilities within the range [0,1], that maximize the constrained secondary throughput. This, however, may hinder the practicality of the suggested algorithm. Hence, in the next subsection we make some observations on the values of the optimization variables under various communication scenarios and use those observations to facilitate a more efficient solution.

### 5.4.1 Simplified Optimization Approach

In order to allow for a more efficient implementation of our algorithm, we suggest here a few measures to provide an approximate solution to the optimization problem in (5.10) with a slight loss of optimality.\[4\]

The approximation steps suggested here are for the cases when $R_p^* > 1 - \delta_p$, i.e. the primary rate requirement is higher than what can be offered by the direct link alone, which is the more
1. First, we note that the objective, i.e. the SU throughput, does not depend on $p_{PS}$, nor does the PU throughput constraint. In fact, the only dependence on $p_{PS}$ in (5.10) is through the stability constraint of the PS queue, which translates to the constraint $p_{PS} > \frac{a_{PS}}{b_{PS}}$ where $b_{PS}^* = b_{PS}/p_{PS}$. This means we can simply choose any $p_{PS}$ value that guarantees the feasibility of our problem. Hence, in the interest of simplifying the algorithm, we fix $p_{PS} = 1$. Note that this step thus far is a simplification, not an approximation.

2. Under the explicit NACK scheme, we make a distinction between two communication scenarios depending on $\Delta_P$.

- Under a large $\Delta_P$ value, concurrent primary and secondary transmissions hurt the PU throughput severely, hence we set $p_N = p_B = 0$. The intuition behind this is that in the explicit NACK case, a NACK is overheard by the SU following a primary transmission failure, which indicates the primary will be active in the next time slot. Also, through sensing, the secondary transmits with probability $p_B$ if the PU is sensed to be busy. Hence, by setting $p_N = p_B = 0$, we can limit the throughput loss due to concurrent transmission. This leaves us with an optimization only over the $\nu$ and $p_F$ values.

- Under a small $\Delta_P$ value, we can simply set $p_F = \nu = 1$ and optimize over the values of $p_B$ and $p_N$. Here, the intuition is that even under the presence of sensing errors, $p_F = 1$ would not hurt the primary transmission much. As for the choice of $\nu = 1$, it was observed from numerical simulation under the small $\Delta_P$ regime.

3. Under the explicit ACK scheme, an ACK state is known with certainty while the interesting communication scenario and the one advantageous for the PU.

\footnote{It may at first appear that $R_S$ depends on $p_{PS}$ through the term $(\psi_0 p_{PS} + \psi_0)$. However, the latter term can be simplified to $1 - p_{PS} \frac{a_{PS}}{b_{PS}}$. Using the expression for $b_{PS}$ in (5.5), it can then be verified that $R_S$ is independent of $p_{PS}$.}
NACK state is ambiguous since it could also indicate inactivity of the primary. Hence, the approximation here is to set $p_F = 1$ and optimize over the remaining values of $\nu$, $p_B$ and $p_N$.

![Figure 5.5](image1)

Figure 5.5: The achievable secondary throughput using the exact and approximate methods under the explicit NACK scheme when $R^*_p = 0.92$.

![Figure 5.6](image2)

Figure 5.6: The achievable secondary throughput using the exact and approximate methods under the explicit ACK scheme when $R^*_p = 0.92$.

Figures 5.5 and 5.6 show the resulting secondary throughput for the explicit NACK and explicit ACK schemes using an exact method versus the approximations suggested above. It is clear that loss of optimality is minimal albeit at a significant gain in computational complexity.

In the following section, we study the queueing delay of the proposed relaying scheme. We then present a slightly modified secondary throughput maximization problem that involves a primary queueing delay constraint.
5.5 Average Queueing Delay for Primary Packets

We start by deriving the delay under the no relaying scenario and then generalize it to the cooperative case. With no cooperation between the primary user and the cognitive one, we apply Little’s law to the PU queue alone. Delay in the PU queue, $D_{PU}$, can be expressed as:

$$D_{PU} = \frac{N_{PU}}{\theta} = \frac{\sum_{k=1}^{\infty} (\pi_k + \eta_k) k}{\theta},$$  \hspace{1cm} (5.11)

where $N_{PU}$ denotes the average number of packets in the PU queue, which we obtain using the steady state probabilities provided in Table 5.1 and is given by

$$N_{PU} = \begin{cases} \frac{\pi_0 \kappa \rho}{(1 - \rho)^2} & \text{for explicit NACK,} \\ \frac{\pi_0 \rho}{(1 - \rho)^2} & \text{for explicit ACK,} \end{cases}$$ \hspace{1cm} (5.12)

with $\kappa = \frac{\theta \beta}{1 - \theta \beta - \theta \gamma}$ and $\rho = \frac{\theta (1 - \theta \beta - \theta \gamma)}{\theta (\theta \beta + \theta \gamma)}$.

On the other hand, under cooperation, the primary packets that are relayed to the ST incur an additional delay in the PS queue and, hence, the queueing delay, $D$, becomes,

$$D = D_{PU} + \epsilon D_{PS},$$ \hspace{1cm} (5.13)

where $D_{PU}$ is given by (5.11), $\epsilon$ is the fraction of primary packets that go to the PS queue and $D_{PS}$ is the delay incurred by such packets. The fraction $\epsilon$ is given by,

$$\epsilon = \frac{\left( \bar{p}_A \left[ \sum_{k=1}^{\infty} \pi_k + \pi_0 \theta \right] + \bar{p}_N \left[ \sum_{k=1}^{\infty} \eta_k + \theta \eta_0 \right] \right) \delta_p \delta_{SC} \nu}{\theta},$$ \hspace{1cm} (5.14)

which is the ratio of the primary packets received successfully by the ST and accepted into the PS queue to the total number of primary packets released from the PU queue,
which either go into the PS queue or directly to the PR.

Applying Little’s law to the PS queue, we derive $D_{PS}$ as,

$$D_{PS} = \frac{N_{PS}}{a_{PS}} = \frac{\sum_{k=1}^{\infty} \psi_k k}{a_{PS}}, \quad (5.15)$$

where $N_{PS} = \frac{a_{PS} \pi_{PS}}{b_{PS}-a_{PS}}$ is the average number of packets in the PS queue. Note that the primary delay may also be derived by applying Little’s law considering the PU and the PS queues as a system, i.e. $D = \frac{N_{PU}+N_{PS}}{\theta}$, which yields an identical result to (5.13).

In order to meet the PU QoS requirements, we now solve the same optimization problem of (5.10) subject to an additional delay constraint, to ensure that the delay does not exceed a specified bound. The delay-constrained optimization problem becomes,

$$\max_{p_{PS}, p_{PU}, p_{PS}, \nu} R_s$$

subject to: $R_p \geq R_p^*$, $D \leq D_{th}$, $\pi_0 > 0$ and $\psi_0 > 0$, \hfill (5.16)

In the numerical analysis section, we compare the secondary throughput attained for both optimization problems. We note here that the same approximation steps that were mentioned in the previous section can be applied here. The only difference is that the dependence on $p_{PS}$ is not merely now in the stability constraint of the PS queue, but also through the delay constraint which cannot be re-written as a linear constraint as in (5.10). However, intuitively, we note that $p_{PS}$ can be still be set to 1 in the interest of minimizing the delay, since once a packet is accepted into the PS queue, a value of $p_{PS} = 1$ will tend to minimize the delay incurred by the primary. In
Figure 5.7: The achievable secondary throughput for various values for the minimum required PU rate, $R_p^*$, under the explicit NACK scheme compared with system simulation verification. The optimal primary packet acceptance probability, $\nu$, and the optimal PS queue selection probability, $p_{PS}$, are 1.

In the numerical simulations, we have also verified that the loss in throughput incurred by the secondary in this case is negligible.

### 5.6 Numerical Results

Here, numerical results are provided for the constrained secondary throughput optimization problems of (5.10) and (5.16). For all links except the secondary’s and the ST-PR link, we choose the transmission failure probabilities as $\delta_i = 0.1$ and $\Delta_i = 0.9$. For the secondary link, we adjust the parameters in equations (5.20) and (5.22) as follows: $L = 1024$ bits, $W = 1$ MHz, $P_S/\sigma_S^2 = 4.903$, $P_P/\sigma_P^2 = 37.9504$, $T = 1 e^{-3}$, $<\gamma_S> = <\gamma_{PS}> = 2$ and $<\gamma_{SP}> = 1$ so that $\delta_S^{(0)} = 0.1$, $\Delta_S^{(0)} = 0.9$ and $\delta_{SP}^{(0)} = 0.1901$, while the values $\delta_S^{(\tau_s)}$, $\Delta_S^{(\tau_s)}$ and $\delta_{SP}^{(\tau_s)}$ vary according to the chosen $\tau_s$ as we show later.

Unless otherwise indicated, the reported results are for the case when $\tau_s$ is considered negligible while $p_{FA}$ and $p_{MD}$ are each set to 0.1. In the last two figures, we report on the achievable SU throughput when the impact of $\tau_s$, with its subsequent modification to the values of $p_{FA}$ and $p_{MD}$, is included.

First, Fig. 5.7 shows the solution of the optimization problem under the explicit NACK scheme for $R_p^* = 0.9, 0.92, 0.94$. Without relaying, only the problem with
\( R^*_p = 0.9 \) is feasible since \( \delta_p = 0.1 \), and in that case this primary rate can only be achieved while the SU remains silent, i.e. yielding a zero secondary throughput, since \( \Delta_p = 0.9 > \delta_p \). Alternatively, the relaying capability allows the PU to achieve rates not possible while using its direct link alone. Also, when the ST cooperates with the PU and acts as a relay, a non-zero secondary throughput can be achieved even when \( R^*_p \geq 1 - \delta_p \). Note that the discontinuities in the analytical results in the figure beyond certain \( \theta \) values stem from the infeasibility of a solution for the given value of \( \theta \) so that \( R_S \) drops to zero. Infeasibility means that no solution exists for the access probabilities that obeys all the constraints of the optimization problem. Further, to validate our queueing analysis, we superimpose on the figures, system simulation results of the achievable secondary throughput using the optimized access probabilities obtained by solving (5.10) offline. The results are averaged over one million runs of the system and clearly validate our approach.

In Fig. 5.8 we plot the corresponding exact access probabilities for Fig. 5.7 for the case when \( R^*_p = 0.94 \). The results validate our approximate approach to set \( p_B = p_N = 0 \) in the case of large \( \Delta_p \) value, hence eliminating a search over these two variables. We note, however, that when all other parameters remain unchanged and this interference drops, i.e. \( \Delta_p = 0.2 \), then the resulting access probabilities are as shown in Fig 5.9. Hence, in this case we set \( p_F = \nu = 1 \) as suggested in the approximate approach.

Fig. 5.10 shows the difference between the explicit ACK and explicit NACK schemes for the relaying case. As noted in [75], under the explicit ACK regime, the absence of primary transmission is interpreted as a NACK. Since the secondary terminal assumes that the PU will transmit, it does not do channel sensing and accesses the channel with probability \( p_N \), which can be zero if the interference is high. On the other hand, when explicit NACK is used, the absence of transmission implies an ACK, which is followed by spectrum sensing. The SU then senses the channel and
with probability $\bar{\theta}$ would find the channel free and transmit with a higher probability than in the former case.

In Fig. 5.11 we plot the difference in secondary throughput between an upper-bound system and our proposed algorithm. The upper-bound system is attained by assuming perfect sensing with negligible sensing time. This leads to the secondary knowing the exact state of the PU at the beginning of each time slot. We note that any POMDP policy with imperfect sensing will lead to a solution somewhere in between our proposed algorithm and the upperbound. Under perfect sensing, the POMDP solution and the upperbound will coincide as the optimal POMDP policy, with perfect knowledge of the primary state, will be myopic, meaning that we will not need to account for the future, since it is enough to make a decision based on
the current state of the primary. Also plotted on this figure is a comparison with an alternative system where sensing only is employed and another that makes use of the ARQ feedback only. As can be seen here, our proposed algorithm provides a clear advantage over those two schemes albeit at a reduced computational complexity than any POMDP solution which lies between our scheme and the upperbound. It is also clear from the figures that relying on the ARQ feedback only leads to a more degradation in performance than relying on the sensing results only. Our goal in presenting these results is to show where our work lies in comparison with existing works in the literature such as \cite{72,73,78,79} albeit different QoS metrics and different settings are considered there.

In Fig. 5.12, we also plot the difference between the proposed scheme and a sensing-only scheme similar to \cite{76}. In that work, optimization is also carried out over the sensing threshold and the secondary power in order to solve the optimization problem of (5.10). This means that we now vary the $p_{FA}$ and $p_{MD}$ probabilities according to the optimal sensing threshold and also vary $\delta_S^{(0)}$, $\delta_S^{(t_s)}$, $\Delta_S^{(0)}$, $\Delta_S^{(t_s)}$, $\delta_{SP}^{(0)}$, $\delta_{SP}^{(t_s)}$ according to the optimal secondary power. As is evident from the curve, our

\footnote{The sensing-only scheme is constructed by considering only two access probabilities, $p_F$ and $p_B$ since these are the only possibilities following the sensing outcome. The ARQ-only system, on the other hand, is constructed considering probability $p_N$ following a NACK, and another probability, $p_A$ following an ACK. In both systems, the throughput formulas are attained by solving a single-class early-arrival Markov chain.}
Figure 5.11: The achievable secondary throughput using the proposed algorithm, an ARQ-only system, a sensing-only system, and an upperbound system, for the explicit NACK scheme, with $R_p^* = 0.9$ and $\Delta_P = 0.2$.

scheme outperforms that of [76] since the ARQ feedback is also incorporated in the analysis.

Figure 5.12: The achievable secondary throughput using the proposed algorithm and a sensing-only system similar to [76] where optimization is also carried out for the sensing threshold and the secondary power for the explicit NACK scheme, with $R_p^* = 0.92$ and $\gamma_{PS} = 0.05$ and $\tau_s = 0.4T$.

In Fig. 5.13 we plot the solution of the optimization problems of (5.10) and (5.16) for various values of $D_{th}$. We note that the impact of the delay-constrained problem is to decrease the achievable secondary throughput while reducing the range of feasibility of the optimization problem.

Finally, in Fig. 5.14 we study the impact of the variation of $\tau_s$ on the achievable secondary throughput. It is clear from Fig. 5.14 that if $\tau_s$ is too small, the reliability of sensing is in question and subsequently the throughput degrades significantly. On
Figure 5.13: The achievable secondary throughput for the constrained and unconstrained delay optimization problems for different values of $D_{th}$. $D_{th} = \infty$ corresponds to the unconstrained delay problem. The minimum guaranteed primary throughput $R^*_p = 0.92$.

the other hand, if $\tau_s$ becomes too large, then the outage probability increases because the SU would still attempt transmission of one packet in the remaining time of the slot, and this would also degrade throughput, hence the sensing-throughput tradeoff we alluded to earlier. Fig. 5.15 shows that the optimal $\tau_s$ value for the chosen system parameters is around $\tau_s = 0.07T$ when we fix $\theta = 0.051$.

Figure 5.14: The achievable secondary throughput for various values of $\tau_s$.

5.7 Conclusions

In this chapter, we developed a probabilistic secondary access mechanism which incorporates the sensing results and the primary ARQ feedback in a spectrum-sharing system whereby a secondary relaying capability is also included for the failed primary
8 packets. In analyzing the resulting system from a queueing theory standpoint, we were able to quantify the throughput and delay of the primary user and the secondary user throughput. We showed that by solving a secondary throughput maximization problem, subject to primary user QoS requirements, the optimal secondary medium access probabilities can be attained. The results clearly quantify the benefits of cognitive relaying to the primary and the secondary users while accounting for practical issues such as sensing time duration.

5.8 Appendix

5.8.1 Outage Probabilities of the secondary link

We derive here, following [87], expressions for the outage probability of the secondary link with and without interference. The other outage probabilities are given by similar expressions using the corresponding parameters. We consider Rayleigh fading channels and assume that the channel gains remain constant over the time slot duration. Assuming that the number of bits in a packet is $L$, the slot duration is $T$ and the transmission duration is $t$, $0 < t \leq T$, the secondary transmission rate is given by

$$r_S = \frac{L}{t}. \quad (5.17)$$
For example, if \( \tau_s \) seconds are used to sense the channel and the rest of the time slot is used for data transmission, \( t = T - \tau_s \). If transmission proceeds directly at the beginning of the time slot, then \( t = T \). Define \( p_{out} \) as the secondary link’s outage probability when the PT is silent. It is equal to \( \delta_S^{(0)} \) when \( t = T \) and to \( \delta_S^{(\tau_s)} \) when \( t = T - \tau_s \). As mentioned in Section 5.2, outage occurs when the transmission rate exceeds the channel capacity. That is,

\[
\delta_S^{(T-t)} = \Pr \left\{ r_S > W \log_2 \left( 1 + \frac{P_S \gamma_S}{\sigma_S^2 W} \right) \right\},
\]

where \( W \) is the bandwidth of the channel, \( P_S \) is the ST’s transmit power, \( \sigma_S^2 \) is the noise power spectral density and \( \gamma_S \) is the gain of the ST-SR link, which is exponentially distributed in the case of Rayleigh fading. Probability \( p_{out} \) can be written as

\[
\delta_S^{(T-t)} = \Pr \left\{ \gamma_S < \frac{2 \pi - 1}{P_S / (\sigma_S^2 W)} \right\}.
\]

Assuming that the mean value of \( \gamma_S \) is \( <\gamma_S> \), then

\[
\delta_S^{(T-t)} = 1 - \exp \left( - \frac{2 \pi - 1}{P_S <\gamma_S> / (\sigma_S^2 W)} \right).
\]

Now we turn our attention to the case of simultaneous primary and secondary transmission. The outage probability in this case is equal to \( \Delta_S^{(0)} \) when \( t = T \) and to \( \Delta_S^{(\tau_s)} \) when \( t = T - \tau_s \). Assume that the PT transmits with power \( P_P \) concurrently with the ST. If the PT-SR channel gain is \( \gamma_{PS} \), the outage probability is given by

\[
\Delta_S^{(T-t)} = \Pr \left\{ r_S > W \log_2 \left( 1 + \frac{P_S \gamma_S}{P_P \gamma_{PS} + \sigma_S^2 W} \right) \right\}.
\]

If \( \gamma_S \) and \( \gamma_{PS} \) are independent and exponentially distributed with mean values \( <\gamma_S> \)
and $<\gamma_{PS}>$, respectively, the outage probability becomes

$$
\Delta_S^{(T-t)} = 1 - \frac{\exp\left(-\frac{2\overline{t}}{P_{S}<\gamma_{PS}>/\sigma^2_S W}\right)}{1 + \left[2\overline{t} - 1\right] P_{P}<\gamma_{PS}>/P_{S}<\gamma_{S}>}.
$$

(5.22)

It is clear from (5.20) and (5.22) that $\Delta_S^{(0)} \geq \delta_S^{(0)}$ and $\Delta_S^{(\tau_s)} \geq \delta_S^{(\tau_s)}$ with equality when $<\gamma_{PS}> = 0$ as in the case of a Z-interference channel.

### 5.8.2 Spectrum Sensing Error Probabilities

The error probabilities depend on the type of spectrum sensor used and on the available information regarding the gain of the sensing channel between PT and ST.

We focus here on the case of an incoherent energy detector which requires only the knowledge of the average channel gain between the PT and the spectrum sensor at the ST. The probabilities $p_{FA}$ and $p_{MD}$ are given by

$$
p_{FA} = 1 - \Gamma_{\text{inc}}\left(\frac{E_T}{\sigma^2_S W}, M\right),
$$

(5.23)

$$
p_{MD} = \Gamma_{\text{inc}}\left(\frac{E_T}{\sigma^2_S W + P_P <\gamma_{PS}>}, M\right),
$$

(5.24)

where $E_T$ is the detection threshold, $M$ is the number of samples taken by the energy detector and $\Gamma_{\text{inc}}$ is the normalized incomplete Gamma function given by

$$
\Gamma_{\text{inc}}(x, a) = \frac{1}{\Gamma(a)} \int_0^x t^{a-1} \exp(-t) dt,
$$

(5.25)

with $\Gamma$ being the Gamma function. If $f_s$ is the sampling frequency of the spectrum sensor, then the sensing duration $\tau_s = M f_s$. 
Chapter 6

Throughput Maximization for Cognitive Radio Networks Using Active Cooperation and Superposition Coding

6.1 Background and Literature Review

One of the popular techniques for the secondary users to gain access to the wireless medium is through relaying the primary signal so that certain primary quality of service requirements such as rate and outage probability constraints are met [76, 77, 89, 90].

Within the realm of cooperative communication, superposition coding, which is introduced in [91], has been known to improve the region of the achievable rates for collaborative wireless users [92]. For example, the works in [93, 94] show two-level superposition coding to be superior to simple decode and forward of the original message. Naturally, because of its benefits, the use of superposition coding was also extended to cognitive relaying [95, 96].
Whether superposition coding is employed or not, the use of cognitive relaying is predicated on the primary user playing an active role in spectrum management. Such role was shown in previous work in the literature to benefit both users [19, 20]. Hence, we also focus in this chapter on a communication scheme whereby the primary users are not oblivious to the presence of the secondary users but play an active and cooperative role to maximize their benefit from this cognitive spectrum access.

Previous works in the literature also considered cooperative primary/cognitive communication. In [89], when the secondary cooperates, time is divided into alternating time slots. In odd slots, the primary transmits its data, while in even slots the secondary transmits a superimposed signal of its message and that of the primary. Only one transmitter is allowed to transmit at a given time. The objective of optimization is to find the optimal secondary power so that the secondary throughput is maximized without degrading primary performance. The primary and secondary time slots are assumed to be fixed in that work. In [90], active cooperation is proposed within a game-theoretic framework so that the users optimize the time durations of primary and secondary communications. The authors find conditions under which cooperation is beneficial to all users in the network, however, both transmitters have fixed power levels. In [93], time is split into two transmission phases wherein the first time phase, the primary user broadcasts its message, while in the second time phase, the secondary user broadcasts its signal superimposed on the primary signal. The authors optimize the time and power allocations so that the secondary throughput is maximized while the primary rate is fixed at the direct link rate. The system resembles the two-hop scheme of [97] which we will investigate in the numerical results section. In [94], on the other hand, a multiple access (MAC) scheme is considered with two primary users sending to a common destination. Network coding and superposition coding are utilized at the cognitive node that superimposes its own message on network-coded primary signals. The use of a MAC channel, however, is different.
from the model we consider here.

In this chapter, we analyze a half-duplex coding scheme for cooperative communication between the primary node and two cognitive nodes to send the primary message to a destination. Through superposition coding, the primary user splits its message into three parts and transmits in two phases, while the secondary users utilize a third transmission phase for their own communication. By actively relaying the primary message in the second transmission phase, the secondary node aims to maximize its benefit from this spectrum access model. We formulate two secondary throughput maximization problems subject to constraints that guarantee to the primary user its rate requirements and the power constraints of all nodes in the network. The optimization variables are the time durations of the communication phases and the primary and secondary power portions.

In the first problem, which we first solved in [98], the secondary nodes pre-allocate a designated portion of their powers to relay the primary message in the second communication phase. By actively optimizing the primary power allocation and the time portions of each transmission phase, we maximize the time left for secondary transmission. In the second throughput problem, we take a different approach and consider not only the time resource left for secondary operation, but also the power resource. By formulating a secondary throughput maximization problem with a total power constraint, we find the solutions of the two optimization problems drastically different from each other, depending on the relative differences between the direct and the relaying channel gains.

The main contributions of this chapter may be summarized as follows: i) We provide analytical solutions for the formulated optimization problems that are validated by a numerical optimization approach; ii) Our work provides a more generalized framework than [89, 90] for the active cooperation between primary and secondary users since we optimize both time durations and powers for the transmission phases;
and iii) Compared to other reference schemes, our scheme is shown to provide better throughput for the secondary users without degrading primary performance. To the contrary, for example, we show instances where our suggested scheme provides more throughput for the primary node than is possible through the direct link alone while achieving a non-zero throughput for the secondary nodes.

The remainder of this chapter is organized as follows. Section 6.2 presents the system model and communication scheme. We formulate and solve the first optimization problem with a partial power constraint in Section 6.3. We solve the optimization problem with a total power constraint in Section 6.4. We present some illustrative numerical results in Section 6.5 along with comparisons with some reference schemes. Finally, we conclude the chapter in Section 8.

6.2 System Model

We consider the cognitive radio network illustrated in Fig. 6.1. The network comprises a primary user (P) seeking to communicate with its destination (D) in the presence of a cognitive pair, the secondary source (S) and receiver (R) who communicate with each other. We consider decode-forward (DF) relaying such that both cognitive nodes (S and R) assist in relaying the primary message in exchange for spectrum access for their own communication.

More specifically, we model a half-duplex channel with time division such that transmission is carried out on three phases as shown in Fig. 6.1. In the first transmission phase of duration $\alpha_1$, the primary broadcasts its message. Both S and R attempt to decode their intended portions of the primary signal while the destination keeps its received message to decode later. In the second transmission phase of duration $\alpha_2$, the primary broadcasts its message and the cognitive nodes broadcast the messages they decode and re-encode to the destination node D. At the end of
this phase, the destination decodes the received messages and this completes the two-phase primary communication. The third transmission phase of duration $1 - \alpha_1 - \alpha_2$, is reserved for the use of the cognitive nodes; i.e. for the communication between terminals S and R.

In this chapter, we concern ourselves with maximizing the throughput of the secondary nodes, i.e. maximizing the length of the third transmission phase and the secondary power remaining for cognitive operation, provided that the primary node is, at least, guaranteed its direct-link rate. To achieve this design objective, the primary and secondary nodes optimize the transmit power and time duration of the first two transmission phases to maximize the time and power resources remaining for cognitive transmission.
6.2.1 Three-Message Superposition Coding Communication Scheme

The considered channel is memoryless and has finite input alphabets \( \mathcal{X} \), \( \mathcal{X}_s \) and \( \mathcal{X}_r \) corresponding to the primary, secondary source and secondary receiver, respectively. The three channel output alphabets are \( \mathcal{Y} \), \( \mathcal{Y}_s \) and \( \mathcal{Y}_r \) corresponding to the destination, secondary source and secondary receiver, respectively. The channel is specified by a conditional distribution \( p(y, y_s, y_r | x, x_s, x_r) \). A \((2^{nR_p}, n)\) code for this channel consists of a message set \( M = 1, 2, ..., 2^{nR_p} \), three encoders (at nodes P, S and R) and three decoders (at D, S and R).

We consider complex Gaussian channels where \( h_{ij} \) denotes the channel gain between node \( i \in \{ p, s, r \} \) and node \( j \in \{ d, s, r \} \). In this subsection, we focus on the communication scheme when \( |h_{ps}| > |h_{pr}| > |h_{pd}| \) to illustrate the basic idea. The extension to other channel conditions will be considered in Section 6.3.

Under such channel circumstances, to send a message \( m \) of rate \( R_p \) to the destination, the primary divides \( m \) into three parts, \( (m_d, m_r, m_s) \) with rates \( R_d \), \( R_r \) and \( R_s \), respectively, where \( R_d + R_r + R_s = R_p \). Message \( m_d \) is decoded at the destination, after the second phase, while message \( m_r \) is intended to be relayed by R. Since \( |h_{ps}| > |h_{pr}| > |h_{pd}| \), message \( (m_r, m_s) \) is intended to be relayed by S, so that the rate at S is \( R_S = R_s + R_r \). Let us denote the codeword as \( x^{\alpha_1n} = [x_1, x_2, ..., x_{\alpha_1n}] \) and \( x^{\alpha_2n} = [x_{\alpha_1n+1}, x_{\alpha_1n+2}, ..., x_{(\alpha_1+\alpha_2)n}] \).

**Codebook Generation**

Fix \( p(u)p(x_r | u)p(l)p(x_s | u, l) p(x | u, l, x_r, x_s) \). Generate \( 2^{nR_r} \) i.i.d. sequences \( u^n(m_r) \sim \prod_{i=1}^{n} p(u_i) \). Then, for each \( u^n \), generate one sequence \( x^n_r(m_r) \sim \prod_{i=1}^{n} p(x_r | u_i) \). Generate \( 2^{nR_s} \) i.i.d. sequences \( l^n(m_s) \sim \prod_{i=1}^{n} p(l_i) \). For each \( l^n \) and each \( u^n \), generate one sequence \( x^n_s(m_s, m_r) \sim \prod_{i=1}^{n} p(x_s | u_i, l_i) \). Finally, we generate for each \( (u(m_r), x_r(m_r), l(m_s), x_s(m_s, m_r)) \) \( 2^{nR_d} \) sequences \( x^n(m_d, m_r, m_s) \) which are distributed...
as $\prod_{i=1}^{n} p(x_i|u_i, x_{ri}, l_i, x_{si})$.

To send a message $m$, node $P$ maps it to the codeword $x^n(m_d, m_r, m_s) \in \mathcal{X}$. During the first phase, $P$ sends $x^{\alpha_1 n}$, while $R$ and $S$ listen. $R$ and $S$ then decode $\tilde{m}_r$ and $(\tilde{m}_s, \tilde{m}_r)$, respectively, from the received signal and re-encode it into the codewords $x^n_r(\tilde{m}_r)$ and $x^n_s(\tilde{m}_s, \tilde{m}_r)$, respectively. During the second phase, $P$ sends $x^{\alpha_2 n}$, $R$ sends $x^{\alpha_2 n}_r$ and $S$ sends $x^{\alpha_2 n}_s$.

Decoding Technique

Under joint typicality decoding, at the end of phase 1, $R$ chooses the unique $m_r$ such that,

$$\left(Y^{\alpha_1 n}_r, U^{\alpha_1 n}(m_r)\right) \in \mathcal{A}_{\epsilon}^{\alpha_1 n}. \quad (6.1)$$

Otherwise, an error is declared. In the above equation, $\mathcal{A}_{\epsilon}^{\alpha_1 n}$ is the set of jointly typical sequences $\{Y^{\alpha_1 n}_r, U^{\alpha_1 n}(m_r)\}$ with respect to the distribution $p(u(m_r), y_r)$, i.e. the set of $\alpha_1 n$-sequences with empirical entropies $\epsilon$-close to the true entropies $[99]$. On the other hand, node $S$ chooses the unique $(m_r, m_s)$ such that,

$$\left(Y^{\alpha_1 n}_s, L^{\alpha_1 n}(m_s), U^{\alpha_1 n}(m_r)\right) \in \mathcal{A}_{\epsilon}^{\alpha_1 n}. \quad (6.2)$$

Otherwise, an error is declared. The destination, on the other hand, applies joint decoding using the signal received during both transmission phases, and chooses the unique $(m_d, m_r, m_s)$ such that

$$\left(Y^{\alpha_1 n}_1, U^{\alpha_1 n}(m_r), L^{\alpha_1 n}(m_s), X^{\alpha_1 n}(m_d, m_r, m_s)\right) \in \mathcal{A}_{\epsilon}^{\alpha_1 n}, \quad (6.3)$$

\footnote{With a slight abuse of notation, we use the same symbol $\mathcal{A}_{\epsilon}$ to denote the different sets of jointly typical sequences where the relevant joint distributions are implicitly understood.}
and

\[
( Y_{2}^{\alpha_{2}n}, U^{\alpha_{2}n}(m_r), L^{\alpha_{2}n}(m_s), X^{\alpha_{2}n}(m_d, m_r, m_s), \]
\[ X^{\alpha_{2}n}(m_r), X^{\alpha_{2}n}(m_s, m_r)) \in A^{\alpha_{2}n}. \]

(6.4)

Otherwise, an error is declared. Accordingly, the rate satisfying the following constraints is achievable.

**Proposition 3**: All rates satisfying

\[
R_{p} \leq \alpha_{1} I( Y_{s}; U, L) + \alpha_{1} I( Y_{1}; X|U, L) + \alpha_{2} I( Y_{2}; X|U, L, X_r, X_s),
\]

(6.5)

\[
R_{p} \leq \alpha_{1} I( Y_{1}; U, L, X) + \alpha_{2} I( Y_{2}; U, L, X, X_r, X_s),
\]

(6.6)

\[
R_{p} \leq \alpha_{1} I( Y_{r}; U) + \alpha_{1} I( Y_{1}; L, X|U) + \alpha_{2} I( Y_{2}; L, X, X_s|U, X_r),
\]

(6.7)

are achievable for some joint distribution \( p(u)p(x_r|u)p(l)p(x_s|u,l)p(x|u,l,x_r,x_s) \)

\( p(y, y_r, y_s|x, x_r, x_s). \)

**Proof.** The proof analogously follows the approach in [100] and can be attained by applying joint typicality arguments to analyze the error events at the decoders.

6.2.2 Communication Scheme for Gaussian Signaling and AWGN Channels

In this subsection, we specialize the rate equations to the case of Gaussian signaling and AWGN channels. The received signals over the two transmission phases, as
depicted in Fig. 6.1 can be written as:

\[ Y_1 = h_{pd} X_1 + Z_1, \]  
\[ Y_r = h_{pr} X_1 + Z_r, \]  
\[ Y_s = h_{ps} X_1 + Z_s, \]  
\[ Y_2 = h_{pd} X_2 + h_{rd} X_r + h_{sd} X_s + Z_2, \]  

where \( X = [X_1^{m_d} X_2^{m_r}] \), \( X_r \) and \( X_s \) are the transmitted signals by P, R and S, respectively, to be described shortly. \( Z_1, Z_2, Z_r \) and \( Z_s \) are independent AWGNs with variance \( \sigma^2 \). Nodes P, R and S have individual power constraints, \( P_p, P_r \) and \( P_s \), respectively. The transmitted signals can be written as:

\[ X_1 = \sqrt{\delta_1 P_p} V + \sqrt{\gamma_1 P_p} U + \sqrt{\mu_1 P_p} L, \]  
\[ X_2 = \sqrt{\delta_2 P_p} V + \sqrt{\gamma_2 P_p} U + \sqrt{\mu_2 P_p} L, \]  
\[ X_r = \sqrt{\gamma_r P_r} U, \]  
\[ X_s = \sqrt{\mu_s P_s} L + \sqrt{\gamma_s P_s} U, \]  

where \( V^n(m_d), U^n(m_r) \) and \( L^n(m_s) \sim \mathcal{N}(0,1) \) are independent. The terms \( \delta_1 \) and \( \delta_2 \) are the portions of P power allocated to \( m_d \) in the first and second phase, respectively, while \( \gamma_1 \) and \( \gamma_2 \) are the power portions of P allocated to \( m_r \), and \( \mu_1 \) and \( \mu_2 \) are the power portions of P allocated to \( m_s \). Also, R uses \( \gamma_r \) portion of its power to forward \( \tilde{m}_r \) to the destination, while S uses \( \mu_s \) to forward \( \hat{m}_s \) and \( \gamma_s \) to forward \( \hat{m}_r \). The primary power portions must satisfy constraints such that,

\[ \alpha_1(\delta_1 + \gamma_1 + \mu_1)P_p + \alpha_2(\delta_2 + \gamma_2 + \mu_2)P_p \leq P_p. \]  

The secondary source and relay also have power constraints, however, we will make distinction between partial and total power constraints in the next two sections.
Corollary 4 : For AWGN channels, all rates satisfying (6.17)-(6.19) are achievable.

\[ R_p \leq \alpha_1 \log \left( 1 + \frac{|h_{ps}|^2 (\gamma_1 + \mu_1) P_p}{\sigma^2 + |h_{ps}|^2 P_p \delta_1} \right) + \alpha_1 \log \left( 1 + \frac{|h_{pd}|^2 \delta_1 P_p}{\sigma^2} \right) \]

\[ + \alpha_2 \log \left( 1 + \frac{|h_{pd}|^2 \delta_2 P_p}{\sigma^2} \right) = I_1, \]  \hfill (6.17)

\[ R_p \leq \alpha_1 \log \left( 1 + \frac{|h_{pd}|^2 P_p (\delta_1 + \gamma_1 + \mu_1)}{\sigma^2} \right) \]

\[ + \alpha_2 \log \left( 1 + \frac{|h_{pd}|^2 P_p (\delta_2 + \gamma_2 + \mu_2) + |h_{rd}|^2 P_r \gamma_r + |h_{sd}|^2 P_s (\mu_s + \gamma_s) + \zeta}{\sigma^2} \right) = I_2, \]  \hfill (6.18)

\[ R_p \leq \alpha_1 \log \left( 1 + \frac{|h_{pr}|^2 P_p \gamma_1}{\sigma^2 + |h_{pr}|^2 P_p (\delta_1 + \mu_1)} \right) + \alpha_1 \log \left( 1 + \frac{|h_{pd}|^2 P_p (\delta_1 + \mu_1)}{\sigma^2} \right) \]

\[ + \alpha_2 \log \left( 1 + \frac{|h_{pd}|^2 P_p (\delta_2 + \mu_2) + |h_{sd}|^2 P_s \mu_s + 2 \sqrt{|h_{pd}|^2 |h_{sd}|^2 \mu_s \mu_2 P_p P_s}}{\sigma^2} \right) = I_3, \]  \hfill (6.19)

where \( \zeta = 2 \sqrt{|h_{rd}|^2 |h_{pd}|^2 \gamma_r \gamma_2 P_r P_p + 2 \sqrt{|h_{rd}|^2 |h_{sd}|^2 \gamma_r \gamma_s P_r P_s} + 2 \sqrt{|h_{pd}|^2 \gamma_s \gamma_2 P_s P_p} + 2 \sqrt{|h_{sd}|^2 |h_{pd}|^2 \mu_s \mu_2 P_p P_s} } \).

**Proof.** The rates are derived by evaluating (6.5), (6.6) and (6.7) for the input in (6.12), (6.13), (6.14) and (6.15). We note that the base 2 of logarithm is omitted for convenience. \( \blacksquare \)
6.3 Communication Scheme For Throughput Maximization with a Partial Power Budget

In this section, we formulate the secondary throughput maximization problem as follows:

$$\min_{\omega} \alpha_1 + \alpha_2,$$

s.t. $$I_1 \geq R_p, \quad I_2 \geq R_p, \quad I_3 \geq R_p,$$

$$R_p \geq \log_2 \left(1 + \frac{|h_{pd}|^2 P_p}{\sigma^2}\right) = R_{dl},$$

$$\alpha_1 (\delta_1 + \gamma_1 + \mu_1) + \alpha_2 (\delta_2 + \gamma_2 + \mu_2) \leq 1,$$

$$\alpha_2 \gamma_r P^p_r \leq P^p_r, \quad \alpha_2 (\mu_s + \gamma_s) P^p_s \leq P^p_s,$$

where $$\omega = [\alpha_1 \alpha_2 \delta_1 \delta_2 \gamma_1 \gamma_2 \mu_1 \mu_2 \gamma_r \gamma_s \mu_s]$$ is the vector of optimization variables.

The objective of minimizing $$\alpha_1 + \alpha_2$$ is equivalent to maximizing the secondary throughput. The first three constraints come from the rate equations of (6.17)-(6.19). The fourth constraint guarantees a minimum rate for the primary user, that is at least equal to the direct link rate, $$R_{dl}$$. The remaining constraints are power constraints, where in addition to the primary power constraint of (7.5), we added partial power constraints for S and R. Note that we consider a partial power constraint wherein S and R each pre-allocate power portions of their total powers, respectively $$P^p_s$$ and $$P^p_r$$, to relay the primary message in the second communication phase. This means, $$P^p_s \leq P_s$$ and $$P^p_r \leq P_r$$, where $$P_s$$ and $$P_r$$ are S and R’s total power budgets, respectively. Next, we state the following proposition for the solution of the above optimization problem.

**Proposition 5**: The primary’s transmitter splits its message according to the channel conditions as follows

1. If $$|h_{ps}| > |h_{pr}| > |h_{pd}|$$, then P splits its message into three parts $$m_d$$, $$m_r$$ and

---

2This constraint can be adjusted to achieve rates higher than the direct link alone as we show in the numerical results section.

3In the next section, we consider the optimization problem with the total power budgets.
as summarized in Algorithm 4.

2. If $|h_{ps}| > |h_{pd}| > |h_{pr}|$, then $P$ splits its message into $m_d$ and $m_s$ only and uses Algorithm 4 with $\gamma_1 = \gamma_s = \gamma_r = 0$.

3. If $|h_{pd}| > |h_{ps}| > |h_{pr}|$ or $|h_{pd}| > |h_{pr}| > |h_{ps}|$, then $P$ sends only $m_d$ through the direct communication link and the cognitive throughput is zero.

4. The case of $|h_{pr}| > |h_{ps}| > |h_{pd}|$ corresponds to case 1 above with the roles of $S$ and $R$ reversed so that $R$ sends $(m_s, m_r)$ while $S$ sends $m_s$ only.

5. The case of $|h_{pr}| > |h_{pd}| > |h_{ps}|$ corresponds to case 2 above with the roles of $S$ and $R$ reversed so that $R$ sends $m_r$ and $S$ is silent.

Proof. See Appendix 6.7.1.

Based on the solution of the optimization problem, we note that for the first transmission case, i.e. when $|h_{ps}| > |h_{pr}| > |h_{pd}|$, we have the transmission phases as follows:

- In phase 1, node $P$ sends $m_r$ and $m_s$ only with powers $\gamma_1^* P_p$ and $\mu_1^* P_p$, respectively. Node $R$ decodes $\hat{m}_r$ and node $S$ decodes $(\hat{m}_r, \hat{m}_s)$.

- In phase 2, node $P$ sends $m_d$ and $m_s$ only with powers $\delta_2^* P_p$ and $\mu_2^* P_p$, respectively, while node $R$ sends $\hat{m}_r$ with power $\gamma_r^* P_p$ and node $S$ sends $(\hat{m}_r, \hat{m}_s)$ with powers $(\gamma_s^* P_p, \mu_s^* P_p)$.

In the numerical results section, we evaluate the performance of the proposed algorithm in comparison with other existing schemes in the literature. Next, we consider a similar optimization problem with total power constraints.
Algorithm 4 Power Allocation for Secondary Throughput Maximization when $|h_{ps}| > |h_{pr}| > |h_{pd}|$

1: $\delta_1^* = \gamma_2^* = 0$
2: for all $\alpha_1 \in [0, 1]$ do
3:   for all $\alpha_2 \in [0, 1]$ do
4:     if $\alpha_1 + \alpha_2 > 1$ then
5:       Break
6:   else
7:     $\gamma_r = 1/\alpha_2$.
8:     Find $\delta_2^*$ such that $f(\delta_2^*) = 0$ where $f(\delta_2)$ is given by (6.47).
9:     if $\delta_2^*$ exists then
10:        Use (6.44), (6.45) and (6.46) to find $\gamma_1^* + \mu_1^*, \mu_2^*$ and $\mu_3^*$.
11:        Use (6.48) to find $\mu_1^*$, then $\gamma_1^* = (6.44) - \mu_1^*$.
12:     else
13:        Search over $\gamma_r \in [0, 1/\alpha_2]$ such that $f(\delta_2^*) = 0$.
14:        if $\delta_2^*$ exists then
15:            Repeat steps 10, 11 above.
16:        else
17:            No feasible solution exists for this $(\alpha_1, \alpha_2)$ pair.
18:        end if
19:     end if
20:   end if
21: end for
22: end for
23: Find the minimum $\alpha_1 + \alpha_2$ such that a feasible solution exists.
6.4 Communication Scheme For Throughput Maximization with a Total Power Budget

In this problem, the cognitive node worries about the total power it will have for all three communication phases, hence the throughput maximization problem is formulated as follows

$$\max_{\omega} (1 - \alpha_1 - \alpha_2) \log_2 \left( 1 + \frac{|h_s|^2 P_s \delta_s}{\sigma^2} \right),$$

s.t. $I_1 \geq R_p$, $I_2 \geq R_p$, $I_3 \geq R_p$,

$$R_p \geq \log_2 \left( 1 + \frac{|h_{ps}|^2 P_p}{\sigma^2} \right),$$

$$\alpha_1(\delta_1 + \gamma_1 + \mu_1) + \alpha_2(\delta_2 + \gamma_2 + \mu_2) \leq 1,$$

$$\alpha_2(\gamma_s + \mu_s) P_s \leq P_s,$$

$$\alpha_2 \gamma_r P_r \leq P_r,$$

where $h_s$ is the channel gain between the secondary source and receiver while $\delta_s = \frac{1 - \alpha_2(\gamma_s + \mu_s)}{1 - \alpha_1 - \alpha_2}$ is the fraction of secondary source power remaining for the third (cognitive) communication phase. By this formulation above, we account for both the time the cognitive user will have for its own communication and the power used to maintain this communication.

**Proposition 6** : The primary’s transmitter splits its message according to the channel conditions as follows

1. If $|h_{ps}| > |h_{pr}| > |h_{pd}|$, then P splits its message into two parts $m_d$ and $m_r$ as summarized in Algorithm 5.

2. If $|h_{ps}| > |h_{pd}| > |h_{pr}|$, then P splits its message into $m_d$ and $m_s$ only and uses Algorithm 6.
3. If \(|h_{pd}| > |h_{ps}| > |h_{pr}|\) or \(|h_{pd}| > |h_{pr}| > |h_{ps}|\), then \(P\) sends only \(m_d\) through the direct communication link and the cognitive throughput is zero.

4. If \(|h_{pr}| > |h_{ps}| > |h_{pd}|\), then \(R\) sends \((m_s, m_r)\) while \(S\) sends \(m_s\) only as summarized in Algorithm 7.

5. The case of \(|h_{pr}| > |h_{pd}| > |h_{ps}|\) corresponds to case 2 above with the roles of \(S\) and \(R\) reversed so that \(R\) sends \(m_r\) and \(S\) is silent.

**Proof.** See Appendix 6.7.2. ■

---

**Algorithm 5** Power Allocation for Secondary Throughput Maximization when \(|h_{ps}| > |h_{pr}| > |h_{pd}|\)

1: \(\delta_1^* = \mu_1^* = \mu_2^* = \mu_s^* = 0, \gamma_r^* = 1/\alpha_2\)
2: *for all* \(\alpha_1 \in [0, 1]\) *do*
3: *for all* \(\alpha_2 \in [0, 1]\) *do*
4: *if* \(\alpha_1 + \alpha_2 > 1\) *then*
5: Break
6: *else*
7: Find \(\delta_2^*\) such that \(l(\delta_2^*) = 0\) where \(l(\delta_2)\) is given by (6.63).
8: *if* \(\delta_2^*\) *exists* *then*
9: Use (6.58), (6.59) and (6.62) to find \(\gamma_1^*, \gamma_2^*,\) and \(\gamma_s^*\).
10: *else*
11: No feasible solution exists for this \((\alpha_1, \alpha_2)\) pair.
12: *end if*
13: *end if*
14: *end for*
15: *end for*
16: Find the minimum \(\alpha_1 + \alpha_2\) such that a feasible solution exists.

Based on the solution of the optimization problem, we note that for the first transmission case, i.e. when \(|h_{ps}| > |h_{pr}| > |h_{pd}|\), we have the transmission phases as follows:

- In phase 1, node \(P\) sends \(m_r\) only with power \(\gamma_1^* P_p\). Node \(R\) decodes \(\tilde{m}_r\) and node \(S\) decodes \(\hat{m}_r\).
Algorithm 6 Power Allocation for Secondary Throughput Maximization when $|h_{ps}| > |h_{pd}| > |h_{pr}|$

1: $\delta_1^* = \gamma_1^* = \gamma_2^* = \gamma_r^* = 0$
2: for all $\alpha_1 \in [0, 1]$ do
3:   for all $\alpha_2 \in [0, 1]$ do
4:     if $\alpha_1 + \alpha_2 > 1$ then
5:       Break
6:     else
7:       Find $\delta_2^*$ such that $f''(\delta_2^*) = 0$ where $f''(\delta_2)$ is given by (6.73).
8:       if $\delta_2^*$ exists then
9:         Use (6.68), (6.69) and (6.70) to find $\mu_1^*$, $\mu_2^*$ and $\mu_s^*$.
10:       else
11:         No feasible solution exists for this $(\alpha_1, \alpha_2)$ pair.
12:     end if
13:   end if
14: end for
15: end for
16: Find the minimum $\alpha_1 + \alpha_2$ such that a feasible solution exists.

Algorithm 7 Power Allocation for Secondary Throughput Maximization when $|h_{pr}| > |h_{ps}| > |h_{pd}|$

1: $\delta_1^* = \mu_2^* = 0$
2: for all $\alpha_1 \in [0, 1]$ do
3:   for all $\alpha_2 \in [0, 1]$ do
4:     if $\alpha_1 + \alpha_2 > 1$ then
5:       Break
6:     else
7:       Find $\delta_2^*$ such that $k(\delta_2^*) = 0$ where $k(\delta_2)$ is given by (6.85).
8:       if $\delta_2^*$ exists then
9:         Use (6.79), (6.80), (6.81), (6.83), and (6.84) to find $\gamma_1^* + \mu_1^*$, $\gamma_2^*$, $\mu_s^*$, $\mu_r^*$ and $\gamma_r^*$.
10:       else
11:         No feasible solution exists for this $(\alpha_1, \alpha_2)$ pair.
12:     end if
13:   end if
14: end for
15: end for
16: Find the minimum $\alpha_1 + \alpha_2$ such that a feasible solution exists.
• In phase 2, node P sends $m_d$ and $m_r$ only with powers $\delta_2^* P_p$ and $\gamma_2^* P_p$, respectively, while node R sends $\tilde{m}_r$ with power $\gamma_r^* P_r$ and node S sends $\hat{m}_r$ with power $\gamma_s^* P_s$.

6.5 Numerical Results

In this section, we show the performance of the proposed schemes under the partial and total power budgets and compare our approach with various reference schemes. To the best of our knowledge, no schemes which employ the same access model that we proposed exist so that a direct comparison is possible. Nevertheless, some of the existing works in the literature can be modified for a fair comparison. We begin with the solution to the partial power budget problem and then proceed to the total power one.

For the optimization problem involving the partial power budget constraint, as a reference scheme, we compare our algorithm to the use of best relay selection [101]. For this comparison, we only consider case 1 of proposition 5, i.e. $|h_{ps}| > |h_{pr}| > |h_{pd}|$.

For best relay selection, the primary picks the node that has the best end-to-end SNR to relay its message. As clear from the simulation parameters shown in Fig. 6.2, the primary picks $R$ as a relaying node in this instance since $\min(|h_{pr}|^2 P_p, |h_{rd}|^2 P_p) > \min(|h_{ps}|^2 P_p, |h_{sd}|^2 P_s)$. We can recover two rate constraints for such scheme by setting $\mu_1 = \mu_2 = \mu_s = 0$ in (6.18) and (6.19). We then solve the throughput optimization problem using a similar approach to the one described in Section 6.3.

Fig. 6.2 shows the attainable secondary throughput using the proposed algorithm, best relay selection and a numerical optimization approach. It is clear from Fig. 6.2 that it is suboptimal to rely only on one link since significant throughput loss can occur. The numerical optimization is carried out using Matlab’s fmincon optimization.

---

4In the coming subsection, we will also consider the throughput averaged over all the possible communication cases.
Figure 6.2: Comparison between proposed algorithm, numerical optimization, and best relay selection schemes. The channel parameters are chosen as follows: $|h_{pd}|^2 = 0.5$, $|h_{pr}|^2 = 1.5$, $|h_{ps}|^2 = 1.7$, $|h_{rd}|^2 = 1.2$, and $|h_{sd}|^2 = 1.6$. The powers are $P_p = 1$, $P_s = 0.5$, and $P_r = 2$ and the noise variance $\sigma^2 = 1$. Based on the transmit power and the channel gains of primary network, the direct link rate $R_{dl} = 0.585$.

...tion function. The solution predicted by numerical optimization is very close to our proposed analytical solution. We note, however, that the numerical solution is very sensitive to the choice of the initial value of the optimization variables and there are instances when it may not converge.

We note that the attainable primary rate is higher than the direct link rate $R_{dl} = 0.585$. This happens since, through the cooperative relaying, we can garner a much higher rate for P than what is possible through the direct link alone while still receiving a non-zero secondary throughput. To gain more insight into our problem, in Fig. 6.3, we plot the optimized $\alpha_1$ and $\alpha_2$ values. It is noted that the most significant gains in throughput for the cognitive network come from minimizing $\alpha_2$, since this is the transmission phase where the cognitive nodes assist primary communication, hence, significant reductions in time duration are possible in this phase.

Now, for the case of the optimization problem with the total power budget constraint, we attempt to compare our work with two reference schemes that are detailed in the next subsection.
Figure 6.3: Figure showing $\alpha_1$ and $\alpha_2$ as a function of the primary rate. The channel parameters are chosen as follows: $|h_{pd}|^2 = 0.5$, $|h_{pr}|^2 = 1.5$, $|h_{ps}|^2 = 1.7$, $|h_{rd}|^2 = 1.2$, and $|h_{sd}|^2 = 1.6$. The powers are $P_p = 1$, $P_s = 0.5$, and $P_r = 2$ and the noise variance $\sigma^2 = 1$. Based on the transmit power and the channel gains of primary network, the direct link rate $R_{dl} = 0.585$.

### 6.5.1 Reference Schemes

Herein, we start by comparing our approach with a two hop scheme which does not utilize the direct link and only uses one relay for transmission as described in [97]. The communication scheme can be summarized as follows: A fraction, $W_1$, of the primary bandwidth is allocated to primary communication, while the remaining spectrum is allocated for cognitive operation. For the secondary user to gain access to this spectrum, it must assist the primary communication in its $W_1$ share of the bandwidth. This assistance is facilitated by dividing the primary time slot into two equal phases such that during the first phase, the primary relays its message to the cognitive node, and during the second phase, the secondary node sends this data to the primary destination. The spectrum released by the primary to the secondary user is optimized by the latter’s ability to meet a target rate for the primary network. If the target rate is not met, no cooperation is possible. While the authors use the direct link rate as the metric for the minimum required primary rate, the direct link itself is not used for communication. The design objective in that work is to maximize the primary user power savings and maximize secondary throughput.

While amplify-and-forward is used in this work, we modified it here to decode-
and-forward, like our scheme. Based on this, the rate equation governing this scheme can be written as

\[ R_p \leq \frac{W_1}{2} \min \left\{ \log_2 \left( 1 + \frac{|h_{ps}|^2 P_{p1}}{\sigma^2} \right), \log_2 \left( 1 + \frac{|h_{sd}|^2 P_{s1}}{\sigma^2} \right) \right\}, \quad (6.22) \]

where \( P_{p1} \) is the power used by the primary over the first transmission phase in the \( W_1 \) bandwidth, i.e. \( \frac{W_1}{2} P_{p1} = P_p \), while \( P_{s1} \) is the fraction of secondary power dedicated to assisting the primary, i.e.

\[ \frac{W_1}{2} P_{s1} + (1 - W_1) P_{s2} = P_s, \quad (6.23) \]

where \( P_{s2} \) is the power used by the secondary for its own operation in the \((1 - W_1)\) bandwidth. It is noted that \( R_p \geq R_{dl} \) should also be satisfied to provide an incentive for the primary to cooperate. In order for \( R_p \) to be maximized in (6.22), we must set \( |h_{ps}|^2 P_{p1} = |h_{sd}|^2 P_{s1} \). Using the previous relations, we are able to obtain the optimal power and bandwidth allocations to allow secondary access. Fig. 6.4 shows the resulting secondary throughput with the variation in the required primary rate.

For fair comparison with the two-hop scheme, since it involves only one relay, we ran our simulations giving this relay the power of the two relays in our model. The results have been averaged over 2000 simulation runs of the Rayleigh distributed channel gains. The superiority of our scheme is clear and we attribute this to a number of key differences from this work’s model: 1) Where possible, we utilize the direct link for communication, 2) we utilize superposition coding which, as argued in the introduction, has been known to improve the region of the achievable rates for collaborative wireless users, and 3) we optimize the duration of the primary user transmission phases which can mean significant throughput gains for the cognitive user without degrading primary performance.

\(^5\)Assume here only S is the node available for relaying.
The second reference scheme that we use takes a different approach to the cognitive access problem. In [102], an interference channel model is assumed between the primary and secondary transmitter-receiver pair, i.e. the sum rate for the primary and secondary users can be written as,

\[
R(P'_s, P'_p) = \log_2 \left( 1 + \frac{P'_p|h_{pd}|^2}{\sigma^2 + P'_s|h_{sd}|^2} \right) + \log_2 \left( 1 + \frac{P'_s|h_s|^2}{\sigma^2 + P'_p|h_{pr}|^2} \right),
\]

(6.24)

where \(P'_p \leq P_p\) and \(P'_s \leq P'_s\) are fractions of the total primary and secondary powers, respectively.\(^6\) For fair comparison, we make the power available to the single relay in the reference scheme \(P'_s = P_s + P_r\). Also, \(h_{pr}\) and \(h_{sd}\) become now interference channels caused by the primary transmitter on the secondary receiver and by the secondary transmitter on the primary receiver, respectively. The authors solve the following optimization problem

\(^6\)In that work, the primary and secondary nodes may opt not to use their full available power for communication. For convenience, we have also used a different notation for the total power of the single relay in the reference scheme as it involves the sum power of the two relays in our scheme.
\[(P'_s, P'_p) = \max \frac{R(P', P'_s, P'_p)}{P'_s, P'_p}, \]
\[s.t. \ 0 \leq P'_s \leq P'_s, \]
\[0 \leq P'_p \leq P_p, \]
\[
\log_2 \left(1 + \frac{P'_p |h_{pd}|^2}{\sigma^2 + P'_s |h_{sd}|^2}\right) \geq R_{\min},
\]
where
\[R_{\min} \leq \log_2 \left(1 + \frac{P'_p |h_{pd}|^2}{\sigma^2}\right) \leq R(0, P_p), \tag{6.26}\]
where \(R_{\min}\) is the minimum required primary rate. Under this interference model, the maximum rate that can be achieved is the direct link rate with the maximum primary power, i.e. \(P_p\), hence the upperbound in \(\text{(6.26)}\), i.e. with respect to our model \(R(0, P_p) = R_{dl}\). The authors show that the maximum sum rate occurs on one of the points that belong to the set
\[S_c = \{(0, P_p), (P'_s, P_p), (P'_s, \eta P'_s + \beta)\}, \tag{6.27}\]
if \(\eta P'_s + \beta \leq P_p\), otherwise
\[S_c = \{(0, P_p), (\frac{P_p - \beta}{\eta}, P_p)\}, \tag{6.28}\]
where \(\eta = (2^{R_{\min}} - 1) |h_{sd}|^2 / |h_{pd}|^2\) while \(\beta = (2^{R_{\min}} - 1) \sigma^2 / |h_{pd}|^2\). In Fig. 6.5 we plot the resulting sum rate against the variation in the direct link channel gain using our approach and the solution suggested in [102]. The first breakpoint in the curve showing the sum rate achievable using our approach comes from the fact that once \(|h_{pd}| > |h_{pr}|\), case 2 of Proposition 3 applies which necessarily degrades throughput since only message \(m_s\) is sent from S and no message is sent from R. For our approach,
cooperation becomes impossible beyond $|h_{pd}| > |h_{ps}|$, $|h_{ps}|^2 = 2$ for this curve, where the second break point occurs, and this is where the only achievable rate is the direct link one. As for the interference model, we purposefully set $|h_{sd}|^2$ to a small value, 0.5, to mimic a small interference case in addition to making the primary rate requirement low, only $0.5R_{dl}$. In that case, we can see that beyond $|h_{pd}| > |h_{ps}|$, the interference model can outperform our scheme. However, under this same small interference value but with a larger rate requirement $R_p = R_{min} = 1.1R_{dl}$, the interference model can only garner the direct link rate while our approach can get more than the direct link rate, in addition to a non-zero throughput for the secondary. This is also clear in Fig. 6.6 where $|h_{sd}|^2 = 0.5$, the rate requirement is small, however $|h_{ps}|^2 = 5$. Under such circumstances, the advantages of our scheme are clear as more throughput is possible through our cooperative model.\footnote{If the primary rate requirement is above the direct link rate, it is not clear what incentives the primary user would have to participate in the model proposed in [102].}
Figure 6.6: Comparison between proposed algorithm and two-hop relaying for the optimization problem involving a total power constraint. The channel parameters are chosen as follows: $|h_{pr}|^2 = 1.5$, $|h_{ps}|^2 = 5$, $|h_{rd}|^2 = 1.2$, and $|h_{sd}|^2 = 0.5$. The powers are $P_p = 1$, $P_s = 0.5$, and $P_r = 2$, $P_t = 2.5$ and the noise variance $\sigma^2 = 1$.

6.6 Conclusion

We considered cooperative communication in a cognitive radio network through a three-message superposition coding scheme. The scheme involves dividing the time resource into three phases whereby the first two phases are used for primary communications while the third phase is used for the cognitive operation. Secondary nodes aim to maximize their throughput while guaranteeing a minimum rate for the primary user. We solved two optimization problems that maximize the aforementioned objective subject to information-theoretic rate constraints and partial and total power constraints. The optimization variables are the time durations of the transmission phases and primary and secondary power allocations to the different message portions. We provided an analytical solution to our optimization problem and showed that significant throughput gains can be achieved for all network users. We also demonstrated the advantage of our proposed scheme compared to a number of reference schemes that include best relay selection, two-hop routing and an interference channel model.
6.7 Appendix

6.7.1 Throughput Maximization Problem with Partial Power Budget

We start by deriving the optimal power allocation vector for throughput maximization for the case of $|h_{ps}| > |h_{pr}| > |h_{pd}|$. As a function of both $\alpha_1, \alpha_2$ and the other variables of $\omega$, the optimization problem is analytically intractable to approach unless a two-dimensional search is carried out over $\alpha_1$ and $\alpha_2$. In this case, for the given $\alpha_1$ and $\alpha_2$ values, the objective becomes a constant, which transforms our problem into a problem of finding a feasible solution that obeys all constraints. Hence, we write the Lagrangian as:

$$ L(\omega) = -\sum_{i=1}^{15} \lambda_i C_i(\omega), \quad (6.29) $$

where $C_1(\omega) = I_1 - R$, $C_2(\omega) = I_2 - R$, $C_3(\omega) = I_3 - R$, $C_4(\omega) = 1 - \alpha_1(\delta_1 + \gamma_1 + \mu_1) - \alpha_2(\delta_2 + \gamma_2 + \mu_2)$, $C_5(\omega) = 1 - \alpha_2\gamma_r$, and $C_6(\omega) = 1 - \alpha_2(\mu_s + \gamma_s)$, $C_7(\omega) = \delta_1$, $C_8(\omega) = \mu_1$, $C_9(\omega) = \gamma_1$, $C_{10}(\omega) = \delta_2$, $C_{11}(\omega) = \mu_2$, $C_{12}(\omega) = \gamma_2$, $C_{13}(\omega) = \gamma_s$, $C_{14}(\omega) = \mu_s$ and $C_{15}(\omega) = \gamma_r$.\footnote{The constraints $C_7$-$C_{15}$ are non-negativity constraints on the optimization variables.} The KKT conditions are:

$$ \Delta_\omega L(\omega^*, \lambda^*) = 0 $$

$$ C_i(\omega^*) \geq 0, \ \lambda_i^* \geq 0, \ \lambda_i^* C_i(\omega^*) = 0. \quad (6.30) $$

For a given $\alpha_1$ and $\alpha_2$, it suffices to find a feasible point that satisfies all the constraints. If many points are found, then they are all equally valid from the point of view of minimizing the constant objective. To this end, we consider activating all the inequality constraints. In this way, we obtain 6 equations, i.e., three rate equations and three power constraints in $9$ variables. Furthermore, differentiating the Lagrangian, we get:
\[
\frac{\partial L}{\partial \delta_1} = -\frac{\lambda_1 g_{pd}}{\sigma^2 + g_{pd} \delta_1 P_p} + \frac{\lambda_1 g_{ps}^2 P_p (\gamma_1 + \mu_1)}{(\sigma^2 + g_{ps} P_p (\delta_1 + \mu_1 + \gamma_1)) (\sigma^2 + g_{ps} P_p \delta_1)} \\
+ \frac{\lambda_3 g_{pr}^2 P_p \gamma_1}{(\sigma^2 + g_{pr} P_p (\delta_1 + \mu_1 + \gamma_1)) (\sigma^2 + g_{pr} P_p (\delta_1 + \mu_1))} - \frac{\lambda_3 g_{pd}}{\sigma^2 + g_{pd} (\delta_1 + \mu_1) P_p} \\
- \frac{\lambda_2 g_{pd}}{\sigma^2 + g_{pd} (\delta_1 + \mu_1 + \gamma_1) P_p} + \frac{\ln 2}{P_p} - \frac{\ln 2}{\alpha_1 P_p} = 0,
\] (6.31)

\[
\frac{\partial L}{\partial \mu_1} = -\frac{\lambda_1 g_{ps}}{\sigma^2 + g_{ps} (\delta_1 + \gamma_1 + \mu_1) P_p} - \frac{\lambda_2 g_{pd}}{\sigma^2 + g_{pd} (\delta_1 + \mu_1 + \gamma_1) P_p} - \frac{\lambda_3 g_{pd}}{\sigma^2 + g_{pd} (\delta_1 + \mu_1) P_p} \\
+ \frac{\lambda_3 g_{ps}^2 P_p (\gamma_1)}{(\sigma^2 + g_{pr} P_p (\delta_1 + \mu_1 + \gamma_1)) (\sigma^2 + g_{pr} P_p (\delta_1 + \mu_1))} + \frac{\ln 2}{P_p} - \frac{\ln 2}{\alpha_1 P_p} = 0,
\] (6.32)

\[
\frac{\partial L}{\partial \gamma_1} = -\frac{\lambda_1 g_{ps}}{\sigma^2 + g_{ps} (\delta_1 + \gamma_1 + \mu_1) P_p} - \frac{\lambda_2 g_{pd}}{\sigma^2 + g_{pd} (\delta_1 + \mu_1 + \gamma_1) P_p} - \frac{\lambda_3 g_{pr}}{\sigma^2 + g_{pr} (\delta_1 + \gamma_1 + \mu_1) P_p} \\
+ \lambda_4 \frac{\ln 2}{P_p} - \lambda_9 \frac{\ln 2}{\alpha_1 P_p} = 0,
\] (6.33)

where \( g_{ij} = |h_{ij}|^2 \) above. We note that for the proposed scheme to be superior to direct transmission, and given the assumption on the channel gains, i.e. \( g_{ps} > g_{pr} > g_{pd} \), then \( \gamma_1 \) and \( \mu_1 \) must be set to non-zero values. Hence, by the KKT conditions, \( \lambda_8 = \lambda_9 = 0 \). Also, since \( \gamma_1 > 0 \) and \( \mu_1 > 0 \), we must have \( \gamma_r > 0 \), \( \mu_s > 0 \) and \( \gamma_s > 0 \),
therefore $\lambda_{13} = \lambda_{14} = \lambda_{15} = 0$. The remaining KKT conditions are

$$\frac{\partial L}{\partial \delta_2} = -\frac{\lambda_1 g_{pd}}{\sigma^2 + g_{pd}P_p\delta_2} - \frac{\lambda_2 g_{pd}}{\omega_{rps}\sigma^2} + \frac{\lambda_4 \ln 2}{P_p} - \frac{\lambda_{10} \ln 2}{\alpha_2 P_p} = 0, \quad (6.34)$$

$$\frac{\partial L}{\partial \mu_2} = -\frac{\lambda_2}{\omega_{rps}} \left[ \frac{g_{pd}P_p}{\sigma^2} + \sqrt{\frac{g_{sd}g_{pd}\mu_2 P_s P_p}{\mu_2}} \right] + \lambda_4 \ln 2 - \frac{\lambda_{11} \ln 2}{\alpha_2} = 0, \quad (6.35)$$

$$\frac{\partial L}{\partial \gamma_2} = -\frac{\lambda_2}{\omega_{rps}} \left[ \frac{g_{pd}P_p}{\sigma^2} + \sqrt{\frac{g_{rd}g_{pd}\gamma_2 P_r P_p}{\gamma_2}} + \sqrt{\frac{g_{sd}g_{pd}\gamma_2 P_s P_p}{\gamma_2}} \right] + \lambda_4 \ln 2 - \frac{\lambda_{12} \ln 2}{\alpha_2} = 0, \quad (6.36)$$

$$\frac{\partial L}{\partial \gamma_r} = -\frac{\lambda_2}{\omega_{rps}} \left[ \frac{g_{rd}P_r}{\sigma^2} + \sqrt{\frac{g_{rd}g_{pd}\gamma_2 P_r P_p}{\gamma_r}} + \sqrt{\frac{g_{sd}g_{rd}\gamma_r P_r P_p}{\gamma_r}} \right] + \lambda_5 \ln 2 = 0, \quad (6.37)$$

$$\frac{\partial L}{\partial \gamma_s} = -\frac{\lambda_2}{\omega_{rps}} \left[ \frac{g_{sd}P_s}{\sigma^2} + \sqrt{\frac{g_{sd}g_{pd}\gamma_2 P_s P_p}{\gamma_s}} + \sqrt{\frac{g_{sd}g_{rd}\gamma_r P_r P_p}{\gamma_s}} \right] + \lambda_6 \ln 2 = 0, \quad (6.38)$$

$$\frac{\partial L}{\partial \mu_s} = -\frac{\lambda_2}{\omega_{rps}} \left[ \frac{g_{sd}P_s}{\sigma^2} + \sqrt{\frac{g_{sd}g_{pd}\mu_2 P_s P_p}{\mu_s}} \right] + \lambda_6 \ln 2 = 0, \quad (6.39)$$

where $\omega_{rps} = 1 + \frac{g_{pd}P_p(\delta_2+\gamma_2+\mu_2)+g_{rd}P_r\gamma_r+g_{sd}P_s(\mu_s+\gamma_s)+\zeta}{\sigma^2}$.

Using (6.38) and (6.39), we get

$$\frac{\lambda_2}{\omega_{rps}} \left( \sqrt{\frac{g_{rd}\gamma_2 P_p}{\gamma_s}} + \sqrt{\frac{g_{rd}\gamma_r P_r}{\gamma_s}} \right) = \frac{\lambda_2}{\omega_{rps}} \sqrt{\frac{g_{pd}\mu_2 P_p}{\mu_s}}. \quad (6.40)$$

Now, we need to check whether $\delta_2 > 0$, $\mu_2 > 0$, $\gamma_2 > 0$, i.e. $\lambda_{10} = \lambda_{11} = \lambda_{12} = 0$. Note that the case of $\mu_2 = 0$ is not possible since it leads to an inconsistency in (6.40). Hence $\mu_2 > 0$ and then $\lambda_{12} = 0$. Now, assume $\gamma_2 > 0$ and the associated Lagrange
multiplier \( \lambda_{12} = 0 \), then from (6.35) and (6.36), we get

\[
\frac{\lambda_2}{\omega_{\text{rps}}} \left( \sqrt{\frac{g_{rd}\gamma_r P_{p}^p}{\gamma_2}} + \sqrt{\frac{g_{sd}\gamma_s P_{s}^p}{\gamma_2}} \right) = \frac{\lambda_2}{\omega_{\text{rps}}} \sqrt{\frac{g_{sd}P_{s}^p}{\mu_2}}.
\]

(6.41)

However, if we compare (6.40) and (6.41), we get by solving the m together the inconsistent result that makes,

\[
\frac{\lambda_2}{\omega_{\text{rps}}} \sqrt{\frac{g_{pd}g_{rd}P_{p}^p P_{p}^r \mu_2 \gamma_r}{g_{sd}P_{s}^p \mu_2}} = -\frac{\lambda_2}{\omega_{\text{rps}}} \sqrt{g_{rd} \gamma_r P_{p}^p}.
\]

(6.42)

If \( \lambda_2 > 0 \), the only resolution to the above dilemma, is that \( \lambda_{11} > 0 \), i.e. \( \gamma_2 = 0 \). Again, in the interest of finding only a feasible solution to our constant objective problem, we choose \( \lambda_2 > 0 \), so that \( \gamma_2 = 0 \), since this choice does not violate any constraints. In this case, also \( \lambda_4 > 0 \), \( \lambda_5 > 0 \) and \( \lambda_6 > 0 \).

Note also that if \( \delta_2 > 0 \), then \( \lambda_{10} = 0 \) and from (6.34) we get

\[
\frac{\lambda_1 g_{pd}}{\sigma^2 + g_{pd} P_{p} \delta_2} = \frac{\lambda_4 \ln 2}{P_{p}} - \frac{\lambda_2 g_{pd}}{\omega_{\text{rps}} \sigma^2} = \frac{\lambda_2}{\omega_{\text{rps}}} \sqrt{\frac{g_{sd} g_{pd} P_{s}^p P_{p}}{\mu_2}},
\]

(6.43)

i.e. \( \lambda_1 > 0 \), hence \( \delta_1 = 0 \). To summarize, we now have the equations of (6.17)-(6.19), (6.40) and the power constraints \( \alpha_1 (\delta_1 + \gamma_1 + \mu_1) + \alpha_2 (\delta_2 + \gamma_2 + \mu_2) = 1 \), \( \alpha_2 \gamma_r P_{p}^p = P_{p}^r \), \( \alpha_2 (\mu_s + \gamma_s) P_{s}^p = P_{s}^p \), to solve for the 7 variables \([\delta_2 \gamma_1 \mu_1 \mu_2 \gamma_r \gamma_s \mu_s]\)\(^9\) As per the primary power constraint (7.5), \( \delta_2 \in [0, 1/\alpha_2] \). Hence, from (6.17), we get

\[
\gamma_1(\delta_2) + \mu_1(\delta_2) = \frac{\sigma^2}{g_{ps} P_{p}} \left[ \frac{R_{p-a_2 \log(1+g_{pd} \delta_2 P_{p})}}{\alpha_1} - 1 \right].
\]

(6.44)

From (7.5), we obtain

\[
\mu_2(\delta_2) = \frac{1 - \alpha_1 (\gamma_1(\delta_2) + \mu_1(\delta_2)) - \alpha_2 \delta_2}{\alpha_2}.
\]

\(^9\)Recall that, in the interest of finding a feasible solution, we satisfy (6.17)-(6.19) with equality.
Using (6.40), we get

$$\mu_s(\delta_2) = \frac{\mu_2(\delta_2)}{\alpha_2 (\eta + \mu_2(\delta_2))},$$  \hspace{1cm} (6.46)$$

where $\eta = \frac{g_{sd} \gamma_r P_p}{g_{pd} P_p}$. Now, substituting the above three equations, along with $\gamma_s(\delta_2) = \frac{1}{\alpha_2} - \mu_s(\delta_2)$, into (6.18), we can numerically find $\delta_2$ such that

$$f(\delta_2) = g_{pd} P_p (\delta_2 + \mu_2(\delta_2)) + g_{sd} (\mu_s(\delta_2) + \gamma_s(\delta_2)) + 2 \sqrt{g_{rd} g_{sd} \gamma_s(\delta_2) \gamma_r P_p^p P_s^p}$$

$$+ 2 \sqrt{g_{sd} g_{pd} \mu_s(\delta_2) \mu_2(\delta_2) P_p^p P_p} - \beta = 0,$$  \hspace{1cm} (6.47)$$

where $\beta = 2 \frac{R_p - \alpha_1 \log \left(1 + \frac{g_{pd} P_p (\gamma_1(\delta_2) + \mu_1(\delta_2))}{\sigma^2}\right)}{\alpha_2} - \frac{g_{sd} \gamma_r P_p}{\sigma^2} - 1$. Once $\delta_2$ is obtained, we have $\mu_s$, $\gamma_2$, $\mu_2$ and the sum $\gamma_1 + \mu_1$. Finally, from (6.19), we have

$$\mu_1 = \frac{\sigma^2 (\xi - 1)}{g_{pd} P_p - \xi g_{pr}},$$  \hspace{1cm} (6.48)$$

where $\xi = \frac{N_0^2/\alpha_1}{\sigma^2 + g_{pr} P_p (\gamma_1 + \mu_1)}$ and $l = R_p - \alpha_2 \log \left(1 + \frac{g_{pd} P_p (\delta_2 + \mu_2) + g_{sd} P_p^p \mu_s + 2 \sqrt{g_{pd} g_{sd} \mu_s P_p P_p^p}}{\sigma^2}\right)$.

We note that it may not be feasible to satisfy all the rate and the power constraints with equality. Specifically when $\gamma_r = 1/\alpha_2$, there may not be a numerical solution for $\delta_2$ in (6.18). In that case, we scan the value of $\gamma_r \in [0, 1/\alpha_2]$ and solve (6.18). If no value for $\delta_2$ can be found, then the solution is not feasible for the chosen value of $\alpha_1$ and $\alpha_2$ and active cooperation is not possible.

Next, we consider the case $|h_{ps}| > |h_{pd}| > |h_{pr}|$. It is not difficult to see that in this case, relying on the secondary receiver is not beneficial for the primary, hence the communication scheme reduces to only using the secondary source as a relay. In the two cases of $|h_{pd}| > |h_{ps}| > |h_{pr}|$ or $|h_{pd}| > |h_{pr}| > |h_{ps}|$, direct transmission is the better option for the primary and no cooperation or throughput gains are possible.
6.7.2 Throughput Maximization Problem with Total Power Budget

The case of $|h_{ps}| > |h_{pr}| > |h_{pd}|$

We want to show that in the interest of maximizing the throughput when we have a total power budget, then $\mu_1 = 0$, i.e. P only splits its message into $m_d$ and $m_r$ and no $m_s$ is sent. Unlike the first throughput problem, here the power constraint for S, being a total power budget, cannot be met with equality, otherwise the secondary throughput drops to zero. Hence, it is required to minimize $\alpha_2$, $\gamma_s$ and $\mu_s$ as much as possible so that more power and time are available for the secondary operation in the third communication phase. Now, consider the rate constraints of $I_1$, $I_2$ and $I_3$. If we assume that all the other parameters are optimized, except for $\gamma_1$ and $\mu_1$, then we find that $I_1$ and $I_2$ both include the sum $\gamma_1 + \mu_1$ and both are increasing functions of the sum $\gamma_1 + \mu_1$ while $I_3$ is a decreasing function of $\mu_1$ and an increasing function of $\gamma_1$.\[10\]

Note also from the primary power constraint, that the sum $\gamma_1 + \mu_1 = 1 - \alpha_1\delta_1 - \alpha_2(\delta_2 + \gamma_2 + \mu_2)$. This means that the highest value of $\gamma_1$, which also corresponds to the lowest value of $\mu_1$, maximizes the log terms in $I_3$ so that this rate constraint can be satisfied at lower $\alpha_1$ values, which means more throughput for the secondary node. Hence, we conclude that $\mu_1 = 0$, in the interest of maximizing the throughput. It follows then that $\mu_s = \mu_2 = 0$.

Hence, our problem can be redefined as

\[10\]Recall that we assume here that $|h_{pr}| > |h_{pd}|$. 
\[
\max_\omega \ (1 - \alpha_1 - \alpha_2) \log_2 \left( 1 + \frac{|h_{\omega}|^2 \delta_s \sigma^2}{\sigma^2} \right),
\]

s.t. \( I'_1 \geq R_p, \ I'_2 \geq R_p, \)

\[
R_p \geq \log_2 \left( 1 + \frac{g_{pd} P_p}{\sigma^2} \right), \tag{6.49}
\]

\[
\alpha_1 (\delta_1 + \gamma_1) + \alpha_2 (\delta_2 + \gamma_2) \leq 1,
\]

\[
\alpha_2 \gamma_s + (1 - \alpha_1 - \alpha_2) \delta_s \leq 1,
\]

\[
\alpha_2 \gamma_r \leq 1,
\]

where \( I'_1 \) and \( I'_2 \) are the updated rate equations that can be obtained by applying joint typicality arguments to analyze the error events at the decoders when P sends only messages \( m_d \) and \( m_r \) similar to the approach used to obtain the results in Proposition 3. The new rate equations are given by

\[
R_p \leq \alpha_1 \min \left\{ \log_2 \left( 1 + \frac{g_{ps} P_p \gamma_1}{\sigma^2 + g_{ps} P_p \delta_1} \right), \log_2 \left( 1 + \frac{g_{pr} P_p \gamma_1}{\sigma^2 + g_{pr} P_p \delta_1} \right) \right\} + \alpha_1 \log \left( 1 + \frac{g_{pd} P_p \delta_1}{\sigma^2} \right)
\]

\[
+ \ \alpha_2 \log \left( 1 + \frac{g_{pd} P_p \delta_2}{\sigma^2} \right) = I'_1, \tag{6.50}
\]

\[
R_p \leq \alpha_1 \log \left( 1 + \frac{g_{pd} P_p (\delta_1 + \gamma_1)}{\sigma^2} \right)
\]

\[
+ \ \alpha_2 \log \left( 1 + \frac{g_{pd} P_p (\delta_2 + \gamma_2) + g_{rd} P_r \gamma_r + g_{sd} P_s \gamma_s + \Gamma_{rps}}{\sigma^2} \right) = I'_2, \tag{6.51}
\]

where \( \Gamma_{rps} = 2 \sqrt{g_{rd} g_{pd} \gamma_r \gamma_2 P_r P_p} + 2 \sqrt{g_{vd} g_{sd} \gamma_r \gamma_s P_r P_s} + 2 \sqrt{g_{sd} g_{pd} \gamma_s \gamma_2 P_s P_p} \). By writing the Lagrangian, we get

\[
\mathcal{L}(\omega) = (1 - \alpha_1 - \alpha_2) \log_2 \left( 1 + \frac{g_s \delta_s}{\sigma^2} \right) + \sum_{i=1}^{5} \lambda_i C_i(\omega), \tag{6.52}
\]

where \( C_1(\omega) = I'_1 - R, \ C_2(\omega) = I'_2 - R, \ C_3(\omega) = 1 - \alpha_1 (\delta_1 + \gamma_1) - \alpha_2 (\delta_2 + \gamma_2), \)
\[ C_4(\omega) = 1 - \alpha_2 \gamma_r \text{, and } C_5(\omega) = 1 - \alpha_2 \gamma_s. \] Differentiating the Lagrangian, we get

\[
\frac{\partial \mathcal{L}}{\partial \delta_1} = -\frac{\lambda_1 g_{pr}^2 P_p \gamma_1}{\sigma^2 + g_{pr} P_p (\delta_1 + \gamma_1)} + \frac{\lambda_2 g_{pd}}{\sigma^2 + g_{pd} P_p \delta_1} - \lambda_3 \frac{\ln 2}{P_p} = 0, \quad (6.53)
\]

\[
\frac{\partial \mathcal{L}}{\partial \gamma_1} = \frac{\lambda_1 g_{pd}}{\sigma^2 + g_{pd} \gamma_1 P_p} + \frac{\lambda_2 g_{pr}}{\sigma^2 + g_{pr} \gamma_1 P_p} - \frac{\lambda_3 \ln 2}{P_p} = 0, \quad (6.54)
\]

\[
\frac{\partial \mathcal{L}}{\partial \delta_2} = \frac{\lambda_1 g_{pd}}{\sigma^2 \omega_{'\text{rps}}} + \frac{\lambda_2 g_{pd}}{\sigma^2 + g_{pd} P_p \delta_2} - \frac{\lambda_3 \ln 2}{P_p} = 0, \quad (6.55)
\]

\[
\frac{\partial \mathcal{L}}{\partial \gamma_2} = \frac{\lambda_1}{\omega_{'\text{rps}}} \left[ \frac{g_{pd}}{\sigma^2} + \sqrt{\frac{g_{rd} g_{pd} \gamma_2 P_p}{\gamma_2 P_p}} + \sqrt{\frac{g_{ss} g_{pd} \gamma_s P_s P_p}{\gamma_s P_p}} \right] - \frac{\lambda_3 \ln 2}{P_p} = 0, \quad (6.56)
\]

\[
\frac{\partial \mathcal{L}}{\partial \gamma_s} = \frac{\lambda_1}{\omega_{'\text{rps}}} \left[ \frac{g_{ss} P_s}{\sigma^2} + \sqrt{\frac{g_{sd} g_{pd} \gamma_2 P_s P_p}{\gamma_s P_s}} + \sqrt{\frac{g_{sd} g_{rd} \gamma_r P_s P_p}{\gamma_s P_s}} \right] - \frac{\lambda_5 \ln 2}{\sigma^2 + g_{ss} P_s (1 - \alpha_1 - \alpha_2)} = 0, \quad (6.57)
\]

where \( \omega_{'\text{rps}} = 1 + \frac{g_{pd}(\delta_2 + \gamma_2) + g_{rd} \gamma_r + g_{sd} \gamma_s + \Gamma_{\text{rps}}}{\sigma^2} \). We note that, similar to the previous problem, \( \delta_1 = 0 \) in order to avoid any inconsistency between equations (6.53) and (6.54). \(^{11}\)

From (6.57), we note that \( \lambda_1 > 0 \), so the first constraint is active, per the KKT conditions. Also, from (6.56), \( \lambda_3 > 0 \), so the primary power constraint is active. Furthermore, using (6.56) in (6.55), it can be shown that \( \lambda_2 > 0 \), and hence the second rate constraint is also active. From the first rate constraint, we have

\[
\gamma_1(\delta_2) = \frac{\sigma^2}{g_{pr} P_p} \left[ 2 \frac{\bar{h}_{pr} - \alpha_2 \log_2 \frac{1 + g_{pd} \delta_2 + \sigma^2}{\alpha_1}}{\bar{h}_{pr}} - 1 \right], \quad (6.58)
\]

\(^{11}\)Assuming \( |h_{ps}| > |h_{pr}| \), then \( \min \left( \log_2 \left( 1 + \frac{|h_{ps}|^2 (\gamma_1 + \mu_1)}{\sigma^2 + |h_{ps}|^2 \delta_1} \right), \log_2 \left( 1 + \frac{|h_{ps}|^2 (\gamma_1 + \mu_1)}{\sigma^2 + |h_{pr}|^2 \delta_1} \right) \right) = \log_2 \left( 1 + \frac{|h_{ps}|^2 (\gamma_1 + \mu_1)}{\sigma^2 + |h_{pr}|^2 \delta_1} \right). \]

\(^{12}\)We also know from the previous throughput problem that in order to maximize the secondary throughput, then \( \gamma_r = 1/\alpha_2 \), so we do not need to investigate the KKT conditions with respect to this variable.
while from the primary power constraint, we get

$$\gamma_2(\delta_2) = \frac{1 - \alpha_1 \gamma_1 - \alpha_2 \gamma_2}{\alpha_2}. \quad (6.59)$$

Using (6.56) in (6.54), we get

$$\lambda_2 = \lambda_1 \frac{\sigma^2 + g_{pd} P_p \delta_2}{\omega'_{rps}} \left[ \sqrt{\frac{g_{rd} P_r \gamma_r}{g_{pd} P_p \gamma_2}} + \sqrt{\frac{g_{sd} P_s \gamma_s}{g_{pd} P_p \gamma_2}} \right]. \quad (6.60)$$

Using (6.60) and (6.56) in (6.54), we get

$$g_{pd} \left[ \frac{1}{\sigma^2 + g_{pd} P_p \gamma_1} - \frac{1}{\sigma^2 \omega'_{rps}} \right] + \left[ \sqrt{\frac{g_{rd} P_r \gamma_r}{g_{pd} P_p \gamma_2}} + \sqrt{\frac{g_{sd} P_s \gamma_s}{g_{pd} P_p \gamma_2}} \right] \left[ \frac{g_{pd}}{\omega'_{rps}} - \frac{g_{pr} (\sigma^2 + g_{pd} P_p \delta_2)}{\omega'_{rps} (\sigma^2 + g_{pr} P_p \gamma_1)} \right] = 0. \quad (6.61)$$

For convenience, let $B = g_{pd} \left[ \frac{1}{\sigma^2 + g_{pd} P_p \gamma_1} - \frac{1}{\sigma^2 \omega'_{rps}} \right]$ and $A = \frac{g_{pd}}{\omega'_{rps}} - \frac{g_{pr} (\sigma^2 + g_{pd} P_p \delta_2)}{\omega'_{rps} (\sigma^2 + g_{pr} P_p \gamma_1)}$.

Then after some manipulations, we arrive at

$$\gamma_s(\delta_2) = \left( \frac{B - A \frac{\sqrt{g_{sd} P_s \gamma_s}}{\frac{g_{pd} P_p \gamma_2}}}{A} \right)^2 \frac{g_{pd} P_p \gamma_2}{g_{sd} P_s}. \quad (6.62)$$

Now, using the values of (6.58), (6.59) and (6.62), we can find a solution to $\delta_2$ from the second rate equation such that

$$l(\delta_2) = \sigma^2 + g_{pd} P_p (\delta_2 + \gamma_2(\delta_2)) + g_{rd} P_r \gamma_r + g_{sd} P_s \gamma_s(\delta_2) + 2 \sqrt{g_{rd} g_{sd} \gamma_s(\delta_2) \gamma_r P_r P_s} + 2 \sqrt{g_{rd} g_{sd} \gamma_2(\delta_2) \gamma_r P_r P_s} + 2 \sqrt{g_{pd} g_{sd} \gamma_2(\delta_2) \gamma_s(\delta_2) P_p P_s} - \sigma^2 \tau = 0, \quad (6.63)$$

where $\tau = 2 \frac{g_{pr} \gamma_1(\delta_2)}{\alpha_2}$. 

The case of $|h_{ps}| > |h_{pd}| > |h_{pr}|$

Under these channel conditions, it is beneficial only to use S for relaying the primary message, hence in this case, only message $m_s$ will be sent. In this case, only two rate constraints, denoted $I_1''$ and $I_2''$, are applicable

\[
R_p = \alpha_1 \log_2 \left( 1 + \frac{g_{ps} \mu_1 P_p}{\sigma^2 + \delta_1 g_{ps}} \right) + \alpha_1 \log_2 \left( 1 + \frac{g_{pd} \delta_1 P_p}{\sigma^2} \right) + \alpha_2 \log_2 \left( 1 + \frac{g_{pd} \delta_2 P_p}{\sigma^2} \right) = I_1'',
\]

\[
R_p = \alpha_1 \log_2 \left( 1 + \frac{g_{pd} (\delta_1 + \mu_1) P_p}{\sigma^2} \right) + \alpha_2 \log_2 \left( 1 + \frac{g_{pd} (\delta_2 + \mu_2) P_p + g_{sd} \mu_s P_s + 2 \sqrt{g_{pd} g_{sd} \mu_s \mu_p P_p P_s}}{\sigma^2} \right) = I_2'',
\]

(6.64)

(6.65)

where $I_1''$ and $I_2''$ can be recovered respectively from $I_1$ and $I_2$, by setting $\gamma_1 = \gamma_2 = \gamma_s = \gamma_r = 0$. Then the throughput problem becomes as follows

\[
\max \omega \ (1 - \alpha_1 - \alpha_2) \log_2 \left( 1 + \frac{g_s \delta_s}{\sigma^2} \right),
\]

\[
s.t. \ I_1'' \geq R_p, \ I_2'' \geq R_p,
\]

\[
R_p \geq \log_2 \left( 1 + \frac{g_{pd} P_p}{\sigma^2} \right),
\]

\[
\alpha_1 (\delta_1 + \mu_1) + \alpha_2 (\delta_2 + \mu_2) \leq 1,
\]

\[
\alpha_2 \mu_s \leq 1,
\]

(6.66)

where $\delta_s = \frac{1 - \alpha_2 \mu_s}{1 - \alpha_1 - \alpha_2}$. By writing the Lagrangian, we get

\[
\mathcal{L}(\omega) = (1 - \alpha_1 - \alpha_2) \log_2 \left( 1 + \frac{|h_s|^2 \delta_s}{\sigma^2} \right) + \sum_{i=1}^{5} \lambda_i C_i(\omega),
\]

(6.67)

where $C_1(\omega) = I_1'' - R$, $C_2(\omega) = I_2'' - R$, $C_3(\omega) = 1 - \alpha_1 (\delta_1 + \mu_1) - \alpha_2 (\delta_2 + \mu_2)$, and $C_4(\omega) = 1 - \alpha_2 \mu_s$. By differentiating the Lagrangian and proceeding in a similar
fashion to the previous problems, it can be shown that the two rate constraints and
the primary power constraint are active and we get an additional equation from the
KKT conditions such that

$$
\mu_1(\delta_2) = \left( \frac{R_p - \alpha_2 \log_2 \left( \frac{\delta_2 g_p P_p}{\sigma^2} \right)}{\alpha_1} \right) \left( \frac{\sigma^2}{g_p P_p} \right) - 1.
$$  \hspace{1cm} (6.68)

$$
\mu_2(\delta_2) = \frac{1 - \alpha_1 \mu_1 - \alpha_2 \delta_2}{\alpha_2},
$$  \hspace{1cm} (6.69)

$$
\mu_s(\delta_2) = \left( \sigma^2 \omega''_{rps} \left[ E + F \right] \right)^2 \frac{\mu_2 P_p}{g_p P_s},
$$  \hspace{1cm} (6.70)

where $E$ and $F$ are defined, respectively, as follows

$$
E = \frac{g_p g_p}{\sigma^2 + \mu_1 P_p g_p} \left[ \frac{1}{\sigma^2 + \mu_1 P_p g_p} - \frac{1}{\sigma^2 + \mu_1 P_p g_p} \right],
$$  \hspace{1cm} (6.71)

$$
F = \frac{g_p}{\sigma^2 + \mu_1 P_p g_p} - \frac{g_p}{\sigma^2 \omega''_{rps}},
$$  \hspace{1cm} (6.72)

where $\omega''_{rps} = 1 + \frac{g_p g_p (\delta_2 + \mu_2) P_p + g_s P_s}{\sigma^2}$.

We can then use these values in the following equation, which results from (6.65), to find the optimal value of $\delta_2$

$$
f''(\delta_2) = g_p P_p (\delta_2 + \mu_2(\delta_2)) + g_s \mu_s(\delta_2) + 2 \sqrt{g_p g_p P_p P_s (\delta_2) P_s P_p - \beta''} = 0,
$$  \hspace{1cm} (6.73)

where $\beta'' = 2 \frac{R_p - \alpha_1 \log_2 \left( \frac{\delta_2 g_p P_p}{\sigma^2} \right)}{\alpha_2} - 1$.

**The case of $|h_{pr}| > |h_{ps}| > |h_{pd}|$**

In this case, R sends $(m_r, m_s)$ while S sends only $m_s$. The rate equations then become
\[
R_p \leq \alpha_1 \log \left( 1 + \frac{g_{pr}(\gamma_1 + \mu_1)P_p}{\sigma^2} \right) + \alpha_1 \log \left( 1 + \frac{g_{pd}\delta_1 P_p}{\sigma^2} \right) + \alpha_2 \log \left( 1 + \frac{g_{pd}\delta_2 P_p}{\sigma^2} \right) = I_1^{m}, \tag{6.74}
\]

\[
R_p \leq \alpha_1 \log \left( 1 + \frac{g_{pd}P_p(\delta_1 + \gamma_1 + \mu_1)}{\sigma^2} \right) + \alpha_2 \log \left( 1 + \frac{g_{pd}P_p(\delta_2 + \gamma_2 + \mu_2) + g_{rd}P_r(\gamma_r + \mu_r) + g_{sd}P_s \mu_s + \zeta''}{\sigma^2} \right) = I_2^{m}, \tag{6.75}
\]

\[
R_p \leq \alpha_1 \log \left( 1 + \frac{g_{ps}P_p\mu_1}{\sigma^2} \right) + \alpha_2 \log \left( 1 + \frac{g_{ps}P_p(\delta_1 + \gamma_1)}{\sigma^2} \right) + \alpha_2 \log \left( 1 + \frac{g_{pd}P_p(\delta_2 + \gamma_2) + g_{rd}P_r\gamma_r + 2\sqrt{g_{pd}g_{rd}\gamma_r\gamma_2 P_p P_r}}{\sigma^2} \right) = I_3^{m}, \tag{6.76}
\]

where \( \zeta'' = 2\sqrt{g_{rd}g_{pd}\gamma_r\gamma_2 P_p P_r} + 2\sqrt{g_{rd}g_{rd}\mu_r\mu_s P_p P_r} + 2\sqrt{g_{sd}g_{pd}\mu_2 P_s P_p} + 2\sqrt{g_{sd}g_{rd}\mu_r^2 P_r P_p} \). So, now the optimization problem becomes

\[
\begin{align*}
\max_\omega \quad & (1 - \alpha_1 - \alpha_2) \log_2 \left( 1 + \frac{|h_s|^2 \delta_s}{\sigma^2} \right), \\
\text{s.t.} \quad & I_1^{m} \geq R_p, \quad I_2^{m} \geq R_p, \quad I_3^{m} \geq R_p, \\
& R_p \geq \log_2 \left( 1 + \frac{g_{pd}P_p}{\sigma^2} \right), \tag{6.78} \\
& \alpha_1(\delta_1 + \gamma_1 + \mu_1) + \alpha_2(\delta_2 + \gamma_2 + \mu_2) \leq 1, \\
& \alpha_2(\gamma_r + \mu_r) \leq 1, \\
& \alpha_2 \mu_s \leq 1.
\end{align*}
\]

Proceeding in a similar fashion to the previous problems, we can show that \( \delta_1 = \mu_2 = 0 \). Also, the first two rate constraints, \( I_1^{m} \) and \( I_2^{m} \) are active, along with the primary and R power constraints, so that
\[
\gamma_1 + \mu_1 = \frac{\sigma^2}{g_{pd} P_p} \left[ 2^{R_p - \alpha_2 \log_2 \left(1 + g_{pd} P_p \delta_2 / \sigma^2 \right)} - 1 \right], \tag{6.79}
\]

\[
\gamma_2 = \frac{1 - \alpha_1 (\gamma_1 + \mu_1) - \alpha_2 \delta_2}{\alpha_2}. \tag{6.80}
\]

By differentiating the Lagrangian and writing the KKT conditions, we get

\[
\gamma_r = \frac{L^2 \gamma_2 P_p}{g_{rd} g_{pd} P_r} \tag{6.81}
\]

where

\[
L = \sigma^2 \omega'_{rps} \left[ \frac{g_{pd}}{\sigma^2 \omega'_{rps} - \sigma^2 + g_{pd} (\gamma_1 + \mu_1) P_p} \right] - 1, \tag{6.82}
\]

where \(\omega'_{rps} = 1 + g_{pd} P_p (\delta_2 + \gamma_2 + \mu_2) + g_{rd} P_p (\gamma_r + \mu_r) + g_{sd} P_s \mu_s + \zeta''\). Furthermore, we have

\[
\mu_s = \frac{\mu_r g_{rd}^2}{g_{sd} L^2 P_s}, \tag{6.83}
\]

and

\[
\frac{g_{pd} \gamma_2 P_p}{\gamma_r} = \frac{g_{sd} \mu_s P_s}{\mu_r}. \tag{6.84}
\]

Using (6.79), (6.80), (6.81), (6.83), and (6.84), we can use the second rate equation to find a solution for \(\delta_2\) such that

\[
k(\delta_2) = \sigma^2 + g_{pd} P_p (\delta_2 + \gamma_2 (\delta_2)) + g_{rd} P_r (\gamma_r (\delta_2) + \mu_r (\delta_2)) + g_{sd} P_s \gamma_s (\delta_2)
+ 2 \sqrt{g_{rd} g_{sd} \mu_s (\delta_2) \mu_r (\delta_2) P_r P_s} + 2 \sqrt{g_{rd} g_{pd} \gamma_2 (\delta_2) \gamma_r (\delta_2) P_p P_r} - \sigma^2 \eta'' = 0, \tag{6.85}
\]

where \(\eta'' = 2 \frac{R_p - \alpha_1 \log \left(1 + g_{pd} P_p (\gamma_1 (\delta_2) + \mu_1 (\delta_2)) / \sigma^2 \right)}{\alpha_2} \).
Chapter 7

User Matching with Relation to the Stable Marriage Problem in Cognitive Radio Networks

7.1 Background and Literature Review

We consider a network of primary and secondary users and focus on a spectrum access approach whereby an SU is allowed to concurrently transmit with a PU. This spectrum access problem was also addressed in [103], where the interaction between the primary and secondary users of the network is modeled as a Stackelberg game. The PU, the leader, jointly determines its power allocation, to guarantee its QoS requirement, and the interference price charged to the SU, to reap revenue. The SU, on the other hand, is the follower and responds to the PU’s decisions by determining its power demands. The game equilibrium is then derived and used to determine the optimization variables. The authors of [103] expanded their model to the single PU-multiple SUs scenario and proposed a heuristic algorithm for the PU to maximize its profit in that case. A question remains as to how to model the interaction of multiple primary and secondary users.
A scenario with multiple PUs and SUs is considered in \[19, 104–110\]. In \[104–109\], PUs lease part of their unused spectrum resources for monetary gains while the SUs buy spectrum opportunities that provide them with the best payoff in terms of performance and price. Rather than a monetary reward, in \[19\], the PUs also lease part of their unused spectrum resources for power saving purposes by using the SUs as relay nodes. In that work, SUs place bids indicating how much power they are willing to spend for relaying the primary signals to their destinations. However, we are interested in the scenario where the PUs require the use of all the available spectrum, hence can only co-exist with the SUs within the same bandwidth, while seeking the monetary benefit of spectrum leasing. Concurrent PU and SU transmission within the same bandwidth is considered in \[110\] where the SUs are allowed to transmit over a number of available subcarriers subject to interference temperature constraints. The transmission strategy of each SU is the power allocation vector over the subcarriers. The authors use the theory of finite variational inequalities to find the solution of the proposed Nash game. A unique Nash equilibrium solution, and convergence to it, is assured only if the interference among the links is sufficiently small.

We use matching theory to study the multiple PUs and SUs case where concurrent transmission of primary and secondary signals is allowed. We propose a three-stage distributed algorithm to pair each PU with an SU. At the first stage, each PU and each SU is considered a potential pair. On this level, we model the interaction between the PU and the SU as a Stackelberg game model and derive the equilibrium utilities each possible pair would get. At the second stage, the derived utilities are sorted to create ordered preference lists for each PU and each SU. In the final stage, the derived preference lists are used to find a stable matching between the PUs and the SUs using the Gale-Shapley algorithm \[111\].

Matching theory and algorithms used to solve the stable marriage problem are relevant to our problem. Within that context, a matching is a one-one pairing of a
set of men to a set of women containing no man and woman who would agree to
leave their assigned partners in order to marry each other \cite{112}. Gale and Shapley
proved that a stable matching always exists and devised an iterative and distributed
algorithm for it \cite{111}.

Matching theory was used within cognitive wireless communication \cite{113,114}. In
\cite{113}, SUs aid PUs by accepting to relay primary packets for a fraction of time,
in exchange for spectrum access. The SUs also pay a monetary compensation to the
PUs. The authors use a utility function which incorporates both the rate and the
monetary factors and devise a matching algorithm whereby PUs and SUs negotiate
the amount of monetary compensation and SU relaying time so that minimum rate
requirements of the PUs and SUs are satisfied. The optimization variables are reached
iteratively.

In \cite{114}, the authors use the Gale-Shapley algorithm to allocate SUs to spectrum
sub-channels. A medium access control technique is introduced to implement the
stable allocation. A consideration of the benefit for or impact on the primary network
from this access is not studied. In fact, the authors consider the utility as the same for
both the SUs and the sub-channels. In that case, there is a unique stable matching.
In our model, however, the PUs and the SUs necessarily receive different utilities from
the postulated spectrum access model. In our case, we are able to show that a unique
stable solution exists. This finding is a direct result of employing the Stackelberg
game model for the interaction between the paired PU and SU.

The contributions of this chapter are as follows: We solve the matching problem for
multiple primary and secondary users in a cognitive network while allowing concurrent
transmission of primary and secondary signals. Our model allows us to study the
interference impact of the cognitive access and to provide the PU with a monetary
compensation for this access while guaranteeing its rate requirement. We solve our
problem via a distributed algorithm which generates a unique stable matching. The
algorithm provides performance close to an optimal centralized controller at a reduced computational complexity. The rest of this chapter is organized as follows. Section 7.2 formulates the problem. Section 7.3 describes the three-stage distributed algorithm. In Section 7.4 we compare the performance of the stable matching to an optimal one and also study the algorithm complexity. Section 7.5 provides some numerical results and Section 7.6 concludes the chapter.

7.2 Problem Formulation

We consider a cognitive network of $N$ PUs and $K$ SUs. The goal of the SUs is to gain access to the $N$ orthogonal primary channels. Only one SU is allowed to coexist with a given PU. SU’s transmission causes interference to the primary receiver, and for this reason an SU is given access to the spectrum in exchange for monetary compensation to the PU. To mitigate interference, the PU also increases its power to meet its rate requirement despite the SU presence. The PU seeks to optimize the interference price it offers to a given SU and its power allocation when paired with this SU to maintain communication. The SU, on the other hand, seeks to optimize its power demand so as to maximize its utility from this spectrum access model. Users, primary or secondary, also seek the pairing that would maximize their respective utilities.

The utility of a given PU, $PU_l$, from allowing access to a given SU, $SU_q$, can be written as,

$$ u_{l,q}^P = \alpha_{l,q} P_{l,q}^S g_{l,q}^{SP} - \beta \left( P_{l,q}^P - P_l^P \right), \tag{7.1} $$

where $\alpha_{l,q}$ is the interference price offered by $PU_l$ to $SU_q$, $P_{l,q}^S$ is the power used by $SU_q$ if paired with $PU_l$ and $g_{l,q}^{SP}$ is the channel gain between $SU_q$’s transmitter and $PU_l$’s receiver, i.e. the interference link to the PU receiver. Parameter $\beta$ is used to

---

1Whether $K$ is greater than $N$ or vice versa is not critical to the distributed algorithm. If $K \neq N$, the the best users will be paired together and some users, whether primary or secondary, will be left unpaired and hence gaining no utility from the matching.
convert the primary power expenditure into a monetary value. Further, $P_{P_{l,q}}^P$ denotes the power that PU$_l$ uses to maintain its minimum rate requirement $r_{P_{l,q}}^P$ if paired with SU$_q$, while $P_{l}^P$ is the power used by PU$_l$ if no access was granted to any SU, i.e. $\log_2 \left( 1 + \frac{P_{P_{l,q}}^P g_{PP_{l,q}}}{n_{l,q}^P} \right) \geq r_{l,q}^P$, where $g_{PP_{l,q}}$ is the primary transmitter to primary receiver channel gain and $n_{l,q}^P$ is the noise power at the receiver of PU$_l$. Based on (7.1), a primary utility matrix is defined: $U^P = < u_{l,q}^P >$.

In addition to maximizing its utility, the PU also requires that any pairing with a given SU must ensure its minimum rate requirement, i.e.,

$$r_{l,q}^P = \log_2 \left( 1 + \frac{P_{P_{l,q}}^P g_{PP_{l,q}}}{P_{l,q} g_{SP_{l,q}} + n_{l,q}^P} \right) \geq r_{l,q}^P.$$  \hfill (7.2)

The utility of SU$_q$ from this spectrum sharing model, on the other hand, can be written as,

$$u_{l,q}^S = \gamma \log \left( 1 + \frac{P_{l,q} g_{SS_{l,q}}}{P_{l,q} g_{SP_{l,q}} + n_{l,q}^S} \right) - \alpha_{l,q} P_{l,q} g_{SP_{l,q}},$$  \hfill (7.3)

where the first term denotes the rate that SU$_q$ gains from the access and $\gamma$ is a constant used to transform that rate into a monetary value. Further, $g_{SS_{l,q}}$ is the channel gain between the secondary transmitter and receiver, $g_{SP_{l,q}}$ is the channel gain between the primary transmitter and the secondary receiver, while $n_{l,q}^S$ is the noise at the receiver of SU$_q$. We define the secondary utility matrix as $U^S = < u_{l,q}^S >$.

In this section, we consider the channel gains to be perfectly known at the beginning of a time slot. Due to the distributed nature of our algorithm, each PU needs only information about the channel gains between itself and the other SUs in the network. In the numerical results section, we remark on the algorithm performance when channel information is known only in a statistical but not exact sense. Further, because the channel gains are continuous random variables, the resulting entries of the utilities matrices, $U^P$ and $U^S$, are almost surely all different. This is an important

\footnote{Without loss of generality (w.l.o.g.), we assume that this value is the same for all PUs.}

\footnote{Hereafter, w.l.o.g. we assume $\gamma = 1$.}
matter in finding a stable matching algorithm as we will discuss later.

Define the primary and secondary user finite and disjoint sets, respectively, as $\mathcal{P} = \{ \text{PU}_i \}_{i=1}^N$ and $\mathcal{S} = \{ \text{SU}_q \}_{q=1}^K$. A matching $M$ is henceforth defined as [115, Ch2, pp.19-20]:

**Definition**: A matching $M$ is a one-to-one correspondence from the set $\mathcal{P} \cup \mathcal{S}$ onto itself of order two (that is, $M^2(x) = x$) such that $M(l) \in \mathcal{S}$ if $M(l) \neq l$ and $M(q) \in \mathcal{P}$ if $M(q) \neq q$. $M(x)$ is referred to as the match of $x$.

In the above definition, $M^2(x) = x$ simply means that if $\text{PU}_i$ is matched to $\text{SU}_q$, i.e. $M(l) = q$, then $\text{SU}_q$ is matched to $\text{PU}_i$, i.e. $M(q) = l$. Also, if $M(l) = l$ then this means that a given user is matched with itself, i.e. with no other member of the opposite set. The definition also requires that PUs and SUs, who are to be paired, be matched with agents of the opposite set. We also define an interference price allocation matrix, $A = \langle \alpha_{l,q} \rangle$, a primary power allocation matrix, $P^p = \langle P^p_{l,q} \rangle$ and a secondary power allocation matrix, $P^s = \langle P^s_{l,q} \rangle$. We note that the elements in $A$, $P^p$ and $P^s$ take the range of values $\{ \alpha_{l,q} \in \mathbb{R} : 0 \leq \alpha_{l,q} \leq \alpha_{\text{max}} \}$, $\{ P^p_{l,q} \in \mathbb{R} : P^p_{l,q} \leq P^p_{l,q, \text{max}} \}$ and $\{ P^s_{l,q} \in \mathbb{R} : 0 \leq P^s_{l,q} \leq P^s_{q, \text{max}} \}$, respectively. The match $M$ and matrices $(A, P^p, P^s)$ are the optimization quantities we seek in order to maximize the utilities of the users in the network.

### 7.3 Distributed Matching Algorithm

The matching $M$ can be obtained by an iterative algorithm similar to the Gale-Shapley algorithm [111]. However, in addition to finding $M$, we have the task of finding $A$, $P^p$ and $P^s$. Hence, we propose the following three-stage algorithm: In the first stage, we use a Stackelberg game to model the interaction between any

---

4In our case this could happen if the user gains zero utility from being matched with any of the users of the opposite set.
matched PU and SU. This framework allows us to solve for the interference price, and the users’ powers, hence the utility matrices \( U^P \) and \( U^S \) are obtained. In the next stage, and in preparation for the matching stage, the utility matrices are sorted to create preference lists for the PUs and the SUs. Finally, in the last stage, we run the classic Gale-Shapley algorithm to find the desired match which we prove to be unique. The remaining subsections describe the three-stage algorithm in detail and discuss measures to ensure protection of the PUs along with an analysis of the algorithm complexity.

### 7.3.1 Stage One: Computing the Utilities

In this stage, the PUs and the SUs compute the equilibrium utilities they would get assuming a Stackelberg game model in which the PU is the leader and the SU is the follower. More specifically, in a Stackelberg game model, the leader, PU\(_l\), makes its decisions first, while SU\(_q\), having observed PU\(_l\)'s decisions, i.e. \( \alpha_{l,q} \) and \( P^P_l \), subsequently makes a decision regarding \( P^S_q \). So, to obtain equilibrium, SU\(_q\) should condition its choice on the observed PU\(_l\) decisions, \( \alpha_{l,q} \) and \( P^P_l \). Hence, the equilibrium solution proceeds by maximizing (7.3) with respect to (w.r.t.) \( P^S_q \), i.e. by setting \( \frac{\partial u^S_{l,q}}{\partial P^S_q} = 0 \). This gives us a value \( P^S_q(\alpha_{l,q}, P^P_l) \). Also, (7.2) can be used to obtain \( P^S_q(P^P_l) \). Now, equating the two obtained values of \( P^S_q \), we obtain \( P^P_l(\alpha_{l,q}) \) and subsequently \( P^S_q(\alpha_{l,q}) \). Using the obtained values, \( u^P_{i,q} \) in (7.1) becomes a function of \( \alpha_{l,q} \) only. By maximizing (7.1) w.r.t. \( \alpha_{l,q} \), we obtain \( \alpha^1_{l,q} \) as,

\[
\alpha^1_{l,q} = \frac{\beta g_{q}^{SS} 2^{P^P_l} - 1}{\ln 2 g_{l,q}^{SP} n_q g_{l}^{PP} + g_{l,q}^{PS} n_l^P \left( 2^{P^P_l} - 1 \right)}.
\] (7.4)
$P_l^P(\alpha_{l,q})$ can be computed from $\alpha_{l,q}$ as,

$$P_l^P(\alpha_{l,q}) = \frac{(2^{r_P} - 1)}{1 + \frac{g_{PS}^{lq} g_{PP}^{lq}}{g_{P}^{lq} g_{SP}^{lq}} (2^{r_P} - 1)} \left\{ \frac{g_{SP}^{lq} 1}{\ln 2 \alpha_{l,q} g_{SP}^{lq} - n_{S}^{q}} - \frac{n_{q}^{S}}{g_{SS}^{q}} + \frac{n_{P}^{l}}{g_{PP}^{lq}} \right\}. \text{ (7.5)}$$

Similarly, $P_q^S(\alpha_{l,q})$ can be computed from $\alpha_{l,q}$ as,

$$P_q^S(\alpha_{l,q}) = \frac{1}{1 + \frac{g_{PS}^{lq} g_{PP}^{lq}}{g_{P}^{lq} g_{SP}^{lq}} (2^{r_P} - 1)} \left\{ \frac{1}{\ln 2 \alpha_{l,q} g_{SP}^{lq}} - \left( \frac{n_{q}^{S}}{g_{SS}^{q}} + \frac{n_{P}^{l}}{g_{PP}^{lq}} \right) \right\}. \text{ (7.6)}$$

In addition to $\alpha_{l,q}^1$, the primary and secondary maximum power constraints, respectively $P_{l,\text{max}}^P$ and $P_{q,\text{max}}^S$, set bounds on the optimal $\alpha_{l,q}$ for the Stackelberg model as follows:

$$\alpha_{l,q}^* = \max\{\alpha_{l,q}^1, \alpha_{l,q}^2, \alpha_{l,q}^3\}, \text{ (7.7)}$$

where $\alpha_{l,q}^*$ denotes the optimal value of the interference price. The value $\alpha_{l,q}^2$ can be computed from (7.3) by setting $P_l^P = P_{l,\text{max}}^P$, while $\alpha_{l,q}^3$ can be computed from (7.6) by setting $P_q^S = P_{q,\text{max}}^S$. Further, in order to guarantee a positive SU power, the following condition must be satisfied:\footnote{This condition can be obtained by setting $P_q^S > 0$ in (7.6).}

$$\beta (2^{r_P} - 1) \left( \frac{n_{q}^{S}}{g_{SS}^{q}} + \frac{(2^{r_P} - 1)n_{P}^{l} g_{SP}^{lq}}{g_{SS}^{q} g_{PP}^{lq}} \right) < g_{P}^{lq} \ln 2 g_{l,q}^{SP}. \text{ (7.8)}$$

Having $\alpha_{l,q}^*, P_l^P(\alpha_{l,q}^*)$ and $P_q^S(\alpha_{l,q}^*)$, we can now calculate the equilibrium primary and secondary utilities to obtain the entries in $U^P$ and $U^S$. Table 7.1 shows a 3 × 3 sample utility matrices. The first entry represents the utility of the PU, while the second entry represents the SU utility. The table shows that the solution is infeasible for the matching of $PU_2$ with $SU_2$, hence they gain a 0 utility if paired, while $PU_3$ and $SU_3$ gain their highest utilities from being matched together. We next build preference lists from the obtained utilities.
Table 7.1: A sample $3 \times 3$ PU and SU utility matrix; $U^P$ entries are in black while $U^S$ entries are in red. For the obtained entries, $P^P_{l,\text{max}} = 2$ and $P^S_{q,\text{max}} = 1$, $\tau^P_l = 2$ for all $l$ and $q$. Channel gains are randomly generated where the average gain for $g^{PP}_{l,q} = g^{SS}_{l,q} = \exp(1)$, while $g^{SP}_{l,q} = g^{PS}_{l,q} = \exp(0.5)$ for all $l$ and $q$.

<table>
<thead>
<tr>
<th></th>
<th>SU$_1$</th>
<th>SU$_2$</th>
<th>SU$_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>PU$_1$</td>
<td>1.08, 1.64</td>
<td>0.15, 0.01</td>
<td>3.6e-3, 2.9e-4</td>
</tr>
<tr>
<td>PU$_2$</td>
<td>0.05, 0.02</td>
<td>0, 0</td>
<td>0.1, 0.04</td>
</tr>
<tr>
<td>PU$_3$</td>
<td>0.6, 0.4</td>
<td>6.3e-3, 5e-4</td>
<td>0.75, 0.7</td>
</tr>
</tbody>
</table>

7.3.2 Stage Two: Creating Preference lists

Let $L^P$ and $L^S$ denote the preference matrices of the PUs and the SUs, respectively. A row $l$ in $L^P$ represents the ordered preferences of PU$_l$ w.r.t. the available SUs, while a column $q$ in $L^S$ represents the ordered preferences of SU$_q$ w.r.t. the available PUs. The preference of PU$_l$ is only for SUs that at least give it a positive utility, i.e. $u^P_{l,q} > 0$. The user topping the list of PU$_l$ will be the one providing the highest PU$_l$ utility from the matching. Similar reasoning applies to the preference criteria of SU$_q$. There may be instances in which a given user labels members of the opposite set as unacceptable, because matching with them gives a zero utility. In this case, the sorting algorithm includes only members of the opposite set that are deemed acceptable.

Table 7.2 shows the resulting $L^P$ and $L^S$ for the matrices in Table 7.1. As expected, PU$_2$ finds SU$_2$ to be unacceptable, and hence the blank entry in the preference row for PU$_2$ (column of SU$_2$).

7.3.3 Stage Three: Gale-Shapley Procedure

Now that we have the ordered preferences of all users in the network, we can proceed to apply the Gale-Shapley algorithm to obtain the matching. In order to appreciate
First, we define a pair that can block a matching.

**Definition**: For a given matching $M$, if there exists a pair $PU_l$ and $SU_q$ who are not matched in $M$, but who prefer each other to their assignments at $M$, then $PU_l$ and $SU_q$ are said to block the matching $M$.

A blocking pair $(l, q)$ means that $P^P(l, q) < P^P(l, M(l))$ coupled with $P^S(l, q) < P^S(M(q), q)$. The presence of any blocking pair in a matching belies its stability as defined next.

**Definition**: A matching $M$ is stable if it is not blocked by any individual or any pair of agents.

Without stability, there is no guarantee that the resulting matching, produced in our case in a distributed fashion, will be enforceable by all parties, specifically if members of the opposite sets have incentives to deviate. With this in mind, we now summarize the steps for the Gale-Shapley algorithm [115, Ch. 2, pp. 27] adapted for our setting.

- At the beginning, all PUs will make a matching proposal to the SU highest on their list. Denoting a matching offer by $\rightarrow$ and using the example of Table 7.1

<table>
<thead>
<tr>
<th>PU</th>
<th>SU1</th>
<th>SU2</th>
<th>SU3</th>
</tr>
</thead>
<tbody>
<tr>
<td>PU1</td>
<td>1, 1</td>
<td>2, 1</td>
<td>3, 3</td>
</tr>
<tr>
<td>PU2</td>
<td>2, 3</td>
<td>-,-</td>
<td>1, 2</td>
</tr>
<tr>
<td>PU3</td>
<td>2, 2</td>
<td>3, 2</td>
<td>1, 1</td>
</tr>
</tbody>
</table>

Table 7.2: $L^P$ (black entries) and $L^S$ (red entries) for the utility matrices in Table 7.1.
this means $PU_1 \rightarrow SU_1$ while $PU_2 \rightarrow SU_3$ and $PU_3 \rightarrow SU_3$. Next, the SUs will in turn look at all matching proposals and pick the one highest on their own preference list. The SU is thus said to be “potentially” matched with this proposing PU. All other PUs are said to be rejected and have to continue looking for a match. SUs not receiving an offer yet are said to be roaming for a match. In our toy example, the end result of this iteration is that $SU_1$ will be potentially matched to $PU_1$, $SU_2$ will be roaming, while $SU_3$ will be potentially matched with $PU_3$.

• On the second iteration of the algorithm, the PUs who have been rejected on the first round will make a matching offer to their second ranking SU (i.e. $PU_2 \rightarrow SU_1$). An SU will then consider the current matching offers, with a potential match if it exists, and choose the one among them highest on its list. All other matches are rejected, i.e. $PU_2$ will again be rejected and now has to remain single since it has no more acceptable partners.

• The algorithm continues until no more matchings can be made. Because the users’ preferences are almost surely strict, the reached matching is stable. In our toy example, we get $M(PU_1) = SU_1$, $M(PU_2) = PU_2$ and $M(PU_3) = SU_3$.

The above procedure can also be seen differently. Consider the maximum entry in $U^P$, we set $M(PU_1) = SU_1$ and delete row 1 and column 1 in $U^P$. Indeed, since $PU_1$ received its first choice, it will no longer consider any other $SU$, while no other $PU_i$ can compete with $PU_1$ on $SU_1$. Now consider the maximum entry in $U^{P-1}$, which is the matrix formed by deleting row 1 and column 1. We can now match $PU_3$ and $SU_3$ and delete their associated rows and columns. Using this procedure, we attain the same matching of the Gale-Shapley procedure above. Moreover, we arrive to the same matching also when we consider the same procedure of iterative deletion of

---

7This is due to the random the channel gains, which makes the entries of $U^P$ and $U^S$, and hence those of $L^P$ and $L^S$, distinct.
rows and columns corresponding to the maximal entries of progressively diminishing matrices but applied to $U^S$. The key observation that enables us to get the same matching is that the indices of the maximal entries of $U^P$ are the same as the indices of the maximal entries of $U^S$, which is a result of the Stackelberg equilibrium as we formalize in the next proposition.

**Proposition 7** The stable matching produced by the three-stage algorithm above is unique.

**Proof.** First, we want to show that the indices corresponding to the maximum entries in $U^P$ are the same as the indices corresponding to the maximum entries in $U^S$. If this true, then it is sufficient to apply the matching procedure to only one of the matrices. We then want to show a unique match exists.

We prove the first part by contradiction. Let $S^P$ and $S^S$ be the vectors containing the elements of $U^P$ and $U^S$ sorted in a descending order, respectively. Let $I^P$ and $I^S$ be the vectors of the indices of $S^P$ and $S^S$ in $U^P$ and $U^S$, respectively. If $I^P \neq I^S$ (component-wise equality), then there must be an entry $S^S(i)$ that is maximized by $\alpha_{l,q}$ and $P^S_{l,q}$ values different from the $\alpha'_{l,q}$ and $P'_{l,q}$ parameters used to maximize $S^P(i)$, but this contradicts the Stackelberg game model assumed for the PU/SU interaction. The second part is a direct result of Proposition V.1. in [114].

### 7.3.4 Protection of Primary Users

We now highlight the measures that we have taken to ensure that the PUs maintain their priority status in the network. First, a PU sets its required rate, as indicated in (7.2). Any matching with a given SU must guarantee this rate. Also, the PU charges the SU for spectrum access. Second, when paired with an SU, the interaction is modeled as a Stackelberg game with the PU being the leader and the SU being the follower. This is advantageous to the PU as it would typically get a higher utility than if, for example, we modeled the interaction as a Cournot game, where both
players move simultaneously. Also, as is clear from Proposition 7, the Stackelberg game modeling ensures that the achieved matching will not only be advantageous to the PU, but also unique. This is a desirable property as it ensures that the SUs will abide by the chosen match. Further, the reached matching possesses a weak Pareto optimality characteristic in the sense that there is no other matching, stable or not, that all PUs prefer to this unique match [115].

Next, we discuss a centralized approach for the purpose of comparison with the proposed algorithm.

### 7.4 A Centralized Controller Optimization Approach

If we had a central controller, then we may use the following two-stage algorithm which upper-bounds our distributed algorithm and, hence, will be called the optimal matching.

- Calculate the utilities $U^P$ and $U^S$ assuming a Stackelberg game model.

- Solve the following optimization problem over all possible matchings:

$$M_{opt} = \arg \max_M u^P(M), \quad (7.9)$$

where $u^P(M)$ is the total primary utility of a matching $M$ defined as $u^P(M) = \sum_{i=1}^{N} u^P_{i,M(i)}$. That is, we find the matching, not necessarily stable, which maximizes the sum utility of the PUs. Although, if we assume for simplicity that there are $N$ PUs and $N$ SUs, there are $N!$ possible matchings, the solution to the optimization problem above can be approached using the Hungarian method, which requires a central controller [114].

In the numerical result section, we compare the outcome of the two approaches, but before showing those results, we discuss the difference in algorithm complexities
where we continue to assume that the number of PUs equals the number of SUs, \( N \).

In stage I (cf. III.A), each PU performs \( O(N) \) operations to calculate the Stackelberg equilibrium over all the SUs. For stage II (cf. III.B), each PU using a typical sorting algorithm can, on average, compute the preference list in \( O(N \log(N)) \) iterations with a worst-case performance of \( O(N^2) \). In stage III (cf. III.C), where the Gale-Shapley algorithm is employed, the algorithm converges to the optimal solution, on average, in \( O(N \log(N)) \) and in a worst-case scenario in \( O(N^2) \). Adding these complexities together, the overall performance of the algorithm is on average \( O(N \log(N)) \), and on a worst-case scenario is \( O(N^2) \). We contrast this with the typical \( O(N^3) \), which is the complexity of the Hungarian method used to solve the optimal assignment problem [114].

### 7.5 Illustrative Numerical Results

We begin by comparing the resulting performance of the proposed distributed algorithm and the optimal one. To simplify the channel gain assignment and to be able to track performance for a large number of users, we use what we term as a *homogeneous assignment* wherein all PUs have the same rate requirement, same direct link channel averages and same interference link averages. We also focus on the case \( N = K \).

We also consider the case when only the channel averages are known. In this case, it can be shown that the expected values of (7.1), (7.2) and (7.3) can be calculated as follows,

\[
\mathbb{E}(u_{i,q}^P) = \alpha_{l,q} P_{t,q}^{SP} - \beta\left(P_{t,q}^P - P_{t,q}^P\right), \tag{7.10}
\]

\[
\mathbb{E}(r_{i,q}^P) = \frac{P_{t,q}^{PP} P_{t,q}^{SP}}{P_{t,q}^{PP} - P_{t,q}^{SP}} \left[ \exp\left(\frac{n_{i,q}^P}{P_{t,q}^{PP}}\right) \psi\left(\frac{n_{i,q}^P}{P_{t,q}^{PP}}\right) - \exp\left(\frac{n_{i,q}^P}{P_{t,q}^{SP}}\right) \psi\left(\frac{n_{i,q}^P}{P_{t,q}^{SP}}\right) \right], \tag{7.11}
\]

\[
\mathbb{E}(u_{i,q}^S) = \frac{P_{t,q}^{SS}}{P_{t,q}^{SS} - P_{t,q}^{SP}} \left[ \exp\left(\frac{n_{i,q}^S}{P_{t,q}^{SS}}\right) \psi\left(\frac{n_{i,q}^S}{P_{t,q}^{SS}}\right) - \exp\left(\frac{n_{i,q}^S}{P_{t,q}^{SP}}\right) \psi\left(\frac{n_{i,q}^S}{P_{t,q}^{SP}}\right) \right] \tag{7.12}
\]
where \( \psi(x) = \int_x^\infty \frac{\exp(-\mu)}{\mu} d\mu \). The Stackelberg equilibrium can be found numerically by first maximizing (7.12) for a given \( \alpha_{l,q} \) and \( P_{l,q}^P \) to obtain \( P_{l,q}^S(\alpha_{l,q}, P_{l,q}^P) \). This value, if it satisfies (7.11), is then used in (7.10) to get the \( \alpha_{l,q} \) and \( P_{l,q}^P \) values. Fig. 7.1 shows a comparison between the stable matching algorithm and the optimal assignment. We also plot the predicted expected utility using the expressions (7.10) and (7.12). The actual expected utility curve is attained by using the expected utility expressions over 10000 channel realizations with the same channel averages. We also plot a random matching algorithm, where we arbitrarily pick a matching from the \( N! \) possible matchings. Note that the random matching we choose requires knowledge of the exact channel gains. The randomness aspect is only over the matching part. The results clearly show the close performance of the proposed stable matching algorithm to the optimal one. It also highlights the superiority of the proposed algorithm to a random assignment. The loss of optimality between the optimal and the distributed algorithm may also be viewed as a stability loss, since the optimal solution, is not guaranteed to yield a stable matching. The results also clearly highlight the loss in utility associated with not knowing the instantaneous SNR and relying on channel statistics only.

Next, we investigate the impact of the interference level on the performance of the proposed algorithm. In Fig. 7.2, we plot the primary sum throughput vs. the number of users for the cases of large interference, where all averages of the interference links are set to 0.5, and small interference, where all interference links averages are set to 0.1. As expected, the sum throughput decreases when we fix all parameters and move from a small to large interference scenario. Further, the difference between the stable and random matching assignments increases as interference increases because the random assignment is more likely to pick a bad match and hence receive a more severe degradation in performance versus the case of small interference.

Finally, we plot in Fig. 7.3 the percentage of matched users vs. the number of
Figure 7.1: A comparison between the stable matching algorithm, the optimal assignment (solved with the Hungarian method) and a random matching assignment for a homogeneous assignment case. All PUs have the same rate requirement (2), same direct link channel averages, set to 1 for both the PP and SS links, and the interference link averages are set to 0.5. Also, $\beta = 1$, $P_{l}^{P,\max} = 2$ and $n_{l}^{P} = 0.1$ for all $l$, $P_{q}^{S,\max} = 1$ and $n_{q}^{S} = 0.1$ for all $q$. For each $N$, results are averaged over 10000 runs of the algorithms.

As $N$ increases, the percentage increases, as there are likely more compatible user pairs. However, this percentage saturates with higher $N$ indicating that this is the maximum percentage of users that can be paired in the network, given the possible utilities of users in the network and the particular channel parameters. As expected, the percentage of paired users is more with the optimal assignment than the stable matching one. We notice, however, that such gap between the two assignments also saturates with increasing $N$ indicating that the loss of throughput between the two schemes seen in Fig. 7.1 can be described again as a price of stability, since the optimal scheme is not guaranteed to achieve a stable matching, and since the proposed matching algorithm necessarily achieves a number of matchings less than or equal to the optimal one. Finally, the percentage of matched users decreases with increasing interference due to the incompatibilities that arise between pairs of users as interference increases.
Figure 7.2: A comparison between the cases of large interference (all parameters are set as in Fig. 7.1) and small interference where the averages of the interference links are set to 0.1.

7.6 Conclusion

We tackled the problem of matching multiple PUs to multiple SUs in a spectrum sharing network. A three-stage distributed and unique stable matching algorithm was proposed to pair users from the two disjoint sets. We showed the superiority of our algorithm to a random pairing of the PUs and the SUs. Compared to an optimal centralized controller, the proposed algorithm was shown to achieve close performance, albeit at a reduced complexity and without sacrificing the stability of the achieved matching. We also quantified the loss in performance associated with the lack/limited knowledge of channel information.
Figure 7.3: A comparison between the percent of users matched under both the stable and optimal matchings for different interference cases.
Chapter 8

Concluding Remarks

8.1 Summary and Future Research Work

In this thesis, we discussed schemes to enhance sensing and channel access in cognitive radio networks. The underlying theme and conclusion of our work is that dynamic spectrum access with active cooperation between the primary and the secondary users can greatly enhance the efficiency of the wireless spectrum and improve the rate and quality of experience for all the involved users.

To summarize, first, we considered the cognitive task of sequentially sensing wide-band primary channels. Under such scenario, we jointly optimized the channel sensing and probing parameters so that the secondary throughput is maximized while the collision probability with the primary network is constrained below a specified threshold. Future extensions of this work would include exploring a computationally efficient algorithm to solve the more challenging problem involving the possibility of multiple secondary users trying to carry out the channel sensing and probing tasks.

Second, we proposed a cooperative spectrum sensing technique to enable coherent combining of sensor decisions at a centralized fusion center despite the sensors’ sending their decisions simultaneously. We considered both soft and hard sensing and compared our equal gain combining approach with a maximum ratio combining one. For all the considered schemes, we derived the global detection probabilities
through a moment generating function solution. We demonstrated the accuracy of our solution through system simulation. We also demonstrated the advantage of the proposed algorithm over other reference and approximate schemes such as the central limit theorem. The robustness of the collaborative sensing scheme is also verified despite practical errors such as phase and synchronization errors. In future extensions of this work, we aim to work on the more general problem of finding the optimal weight vector to use at each sensor, which is not necessarily the equal gain combining solution, so that either the global detection performance is improved or the secondary throughput is maximized given constraints that guarantee quality of service parameters for the primary.

Third, in spectrum access, we optimized a cognitive user scheduling policy as to minimize the cognitive delay under constraints on the average transmit power and the maximum tolerable primary outage probability caused by miss-detection. We considered delay-sensitive and delay-tolerant schemes with respect to the secondary’s reaction to transmission errors. For all considered scenarios, although the delay minimization problem is non-convex, we showed that the access policies can be obtained efficiently using linear programming and grid search over one or two parameters. Future work would involve the exact analysis of the interaction between the primary and secondary queues. It may also tackle the assignment problem of matching the primary and secondary users in a multichannel system. Also of practical interest is a multi-rate system with both users capable of adapting their transmission rates on the basis of their respective channels.

Fourth, we considered the incorporation of both the primary ARQ and the spectrum sensing results in the access decisions of the cognitive network. We obtained the optimal cognitive scheduling probabilities via maximizing the secondary rate subject to primary user rate and delay constraints. Cognitive relaying was also used in this work to aid primary transmission and to offset any disturbance to the primary
network due to the cognitive operation. Our results highlight the benefit of using the ARQ information and of using the relaying scheme with its ability to garner throughput for both the primary and the secondary user. The scenario of ternary feedback, where an ACK is sent when the primary packet is received correctly, an NACK when it is received erroneously and nothing is sent in the case of primary inactivity, is not covered in this thesis and could be part of future extension of this topic.

Fifth, we proposed a three-message superposition coding scheme. The scheme works by dividing the primary time resource into three transmission phases whereby cognitive relaying is used in the second transmission phase, while the third phase is reserved for secondary operation. We formulated two constrained secondary throughput maximization problems. The subject of optimization is the primary and the secondary power portions dedicated to each superposed message and the time durations of the different transmission phases. We derived analytical solutions to both problems under various conditions on the direct and relaying channels. We also demonstrated the advantages of the proposed method against existing works in the literature. In future extensions of this work, we consider the use of another objective for the primary and secondary networks, namely the minimization of total energy, and the impact this may have on the solution of the optimization problem.

Sixth, we proposed a three-stage distributed matching algorithm to pair the primary and secondary users of a cognitive network. A Stackelberg game model is assumed for the interaction between the paired primary and secondary users so that the selected secondary user is given access in exchange for monetary compensation to the primary. We showed a unique stable matching is produced by our algorithm and we also showed its advantages to other reference schemes. More work is needed on understanding how the algorithm would perform if exact channel information is not known.
REFERENCES


Publications

Journal Papers

To be submitted


To appear


Published


**Conference Papers**

