COGNITIVE COMMUNICATIONS FOR EMERGING WIRELESS SYSTEMS

A Dissertation

Presented to

The Graduate Faculty of The University of Akron

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

Ardalan Alizadeh

August, 2016
ABSTRACT

The current explosion of information and demand for high speed data communication call for novel solutions to utilize the radio resources more efficiently. The cognitive communication paradigm aims to mitigate this spectrum crunch by exploiting unused resources that are allocated to the primary communications systems. The aim of this research is to employ the concept of cognition in wireless devices and combine it with three recently introduced wireless communication techniques namely, $K$-user multi-input multi-output (MIMO) interference networks, spatial modulation scheme and molecular communications. Firstly, the feasibility of cognitive radio (CR) is studied in the presence of a $K$-user MIMO interference channel as the primary network. Assuming that the primary interference network has unused spatial degrees of freedom, the sufficient condition on the number of antennas is investigated at the secondary transmitter under which the secondary system can communicate and then the secondary precoding and decoding matrices are derived to have zero interference leakage into the primary network. A fast sensing method based on the eigenvalue analysis of the received signal covariance matrix is proposed to determine the availability of spatial holes. Also, a fine sensing method is provided based on the generalized likelihood ratio test to decide the absence of individual primary streams. The second
part of this research is relevant to the application of \textit{spatial modulation (SM) in overlay CR networks}, in which the primary and secondary networks work concurrently over the same spectrum band. The CR transmitter assists the primary network as a relay to amplify-and-forward (AF) the transmitted symbols of the primary. The secondary transmitter retransmits the primary symbols in amplitude-phase modulation domain, while its own information is transmitted by the index of transmitting antenna. The performance of the optimal detectors in terms of the average symbol error rate (ASER) and the asymptotic behavior of the ASER at both the primary and secondary at high signal-to-noise ratios (SNRs) are then provided. In the last part, a novel \textit{nanonetwork with cognitive capabilities} is proposed, which is able to intelligently sense a primary molecular channel and use the channel opportunistically for its own transmission. In the proposed method, the secondary nanonetwork measures the concentration of molecules as a criterion to decide the presence or absence of the primary communication using a molecular energy detection scheme. When the molecular channel is available, the secondary transmitter sends its information using the same carrier molecules. Depending on the availability of timing information at the sensing nanodevice, two synchronous and asynchronous sensing mechanisms are provided based on the likelihood ratio test as the optimal detection method.
ACKNOWLEDGEMENTS

This work would not have been possible without the support of many people who encouraged and helped me during my doctoral program. I was fortunate to have one of the best advisors both in his scientific abilities and, perhaps more importantly, in being a caring human being. So, I wish to express my gratitude to my advisor Dr. Hamid Reza Bahrami, for his continuous support during my Ph.D. study and research. He supported me through the challenging, and sometimes stressful, situations of my lifetime. His positive attitude, understanding, caring, and immense knowledge have helped me a lot to shape my professional career. I would like to thank and acknowledge my dissertation committee members, Prof. Nathan Ida, Dr. Nghi Tran, Prof. Ping Yi, and Dr. Malena Espanol, for all of their guidance and constructive comments throughout this process.

I thank my family, especially my mom, dad, and brothers Mehran and Esfandiar who have educated me with lots of affection and taught me to be who I am. They did whatever they could to help me through my scientific (and non-scientific) endeavors. The same is true for my wife’s family, especially for her very supporting and kind parents, Ghodratollah and Mehri.

I would also like to thank my many amazing friends all around the world
for helping me during this journey and for creating so many unforgettable memories.

From the long list of friends, I especially like to thank Mehdi Maleki, Mohammed Eltayeb, Mehdi Sadeghzadeh, Ahmad Danaee and Ehsan Saeedpour from Wireless Communications Laboratory (WCL), and Mohammad Moghadamfalahi, Farzad Ahmadi, Mohammad Ranjbar, Nina Moradi, Michael and Mariko Becker, Dr. Sima Chegini, Mehdi Mokhtari, Sam Mahmoudi, Hamid Shafae, Navid Helmi, Kiarash Koushfar, and Amir Tadayon.

I acknowledge and appreciate the Department of Electrical and Computer Engineering for offering me numerous teaching assistantships, Dean’s Summer Fellowship and my advisor for his financial support as research assistantship at WCL.

Finally, and most importantly, I wish to thank very carefully the most important people in my life: my girlfriend, wife, fiancé and best friend, Ronak Moradi. I could never have accomplished this dissertation without her love, support, and encouragement. You are the most beautiful person that I have ever met and I can not explain how much important and precious you are to me. I know you have given up many things to be at my side during all these years. This is why I dedicate this work to you.
## TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>LIST OF TABLES</th>
<th>ix</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF FIGURES</td>
<td>x</td>
</tr>
</tbody>
</table>

### CHAPTER

**I. INTRODUCTION**

1.1 Background .................................................. 1

1.2 Research Objective and Outlines ......................... 4

1.3 Literature Survey ........................................... 11

**II. COGNITIVE RADIO COMMUNICATION IN THE PRESENCE OF A K-USER INTERFERENCE PRIMARY NETWORK**

2.1 System Model ................................................. 18

2.2 Secondary Interference Alignment ......................... 24

2.3 Null Space Sensing and Detection ......................... 32

2.4 Search for the Index Set of Unused DoFs ................. 38

2.5 Simulation Results of Cognitive Interference Alignment Networks ................................ 48

**III. COGNITIVE SPATIAL MODULATION**

3.1 System Model ................................................. 57

3.2 Optimal Detection ........................................... 62

vii
3.3 Average Symbol Error Rate ............................................ 65
3.4 Asymptotic Analysis of ASER ........................................... 74
3.5 System Design .............................................................. 76
3.6 Simulation Results .......................................................... 78
IV. COGNITIVE MOLECULAR COMMUNICATIONS ....................... 86
  4.1 System Model and Problem Formulation ............................... 87
  4.2 Synchronous Molecular Channel Sensing ............................. 96
  4.3 Asynchronous Molecular Channel Sensing ........................... 105
  4.4 Numerical Results ....................................................... 111
V. SUMMARY AND FUTURE RESEARCH .................................. 117
  5.1 Conclusion ................................................................. 117
  5.2 Future Research .......................................................... 119
APPENDICES ................................................................. 122
  APPENDIX A. PROOF OF LEMMA 3 ..................................... 123
  APPENDIX B. PROOF OF LEMMA 5 ..................................... 125
  APPENDIX C. PROOF OF THEOREM 1 ................................. 126
  APPENDIX D. PROOF OF LEMMA 6 ..................................... 127
  APPENDIX E. PROOF OF LEMMA 8 ..................................... 128
BIBLIOGRAPHY ............................................................... 130
### LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Maximum number of antennas at the secondary transmitter for different primary APM sizes and error thresholds, unit channel variances and $P_t/N_0 = 50$ dB.</td>
<td>75</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>An opportunistic cognitive radio system in the presence of a $K$-user interference alignment primary network. When the $i^{th}$ primary stream is inactive, the secondary system is able to utilize the unused DoFs for its own transmission.</td>
<td>19</td>
</tr>
<tr>
<td>2.2</td>
<td>A secondary system in the presence of a primary 3-user IA system; left: all the primary transmitters are active and there is no unused DoFs. right: the second primary transmitter is off and there is one unused DoF that can be used by the secondary system.</td>
<td>26</td>
</tr>
<tr>
<td>2.3</td>
<td>Interference power at the primary and secondary receivers for different scenarios when there is one inactive primary user.</td>
<td>50</td>
</tr>
<tr>
<td>2.4</td>
<td>Interference power at the primary and secondary receivers for different scenarios when there are two inactive primary users.</td>
<td>51</td>
</tr>
<tr>
<td>2.5</td>
<td>Average sum-rate of the primary system versus SNR before and after CR transmission when there is one inactive primary user. Figure also shows the average rate of the secondary for different antenna configurations.</td>
<td>52</td>
</tr>
<tr>
<td>2.6</td>
<td>Average sum-rate of the primary system versus SNR before and after CR transmission when there are two inactive primary users. Figure also shows the average rate of the secondary for different antenna configurations.</td>
<td>53</td>
</tr>
<tr>
<td>2.7</td>
<td>Average sum-rate of the primary system versus number of secondary receive antennas before and after CR transmission, SNR = 15dB and $M^{(0)} = 3$.</td>
<td>54</td>
</tr>
<tr>
<td>2.8</td>
<td>PFA, its lower and upper bounds, and the PD versus threshold for null space sensing.</td>
<td>55</td>
</tr>
</tbody>
</table>
2.9 ROC curves of the null space sensing method for different smoothing factors and number of secondary receive antennas. 56
2.10 ROC curves of the DoF index search method for different smoothing factors and number of secondary receive antennas. 56
3.1 System model of the proposed SM-based overlay CR scheme. 60
3.2 SER of the primary as a function of primary APM size and secondary antenna size, unit channel variances and $P_t/N_0 = 40$ dB. 78
3.3 SER of the primary system for different sizes of APM and spatial domains. 80
3.4 ASERs of the secondary system for different sizes of APM and spatial domains. 81
3.5 ASERs of the primary and secondary systems versus number of secondary antennas ($N_R$) for different APM sizes; $P_t/N_0 = 30$dB. 82
3.6 ASERs of the primary and secondary systems with multi-antenna receivers, $N_R = M = 4$. 84
3.7 Performance comparison of the proposed overlay scheme and an underlay CR scheme, $N_R = M = 4$. 85
4.1 A pair of cognitive nanomachines in the vicinity of a primary nanonetwork with paracrine signaling. 89
4.2 A pair of cognitive nanomachines in the presence of a primary nanonetwork with endocrine signaling. 91
4.3 The probability of arrival of one molecule at the distance of $r_s = 30 \mu$m versus velocity of medium, $D = 1 \text{nm}^2/\text{ns}$. 94
4.4 The sensing timing scheme for three different methods. 97
4.5 The ROC of the synchronous single-pulse sensing with simple thresholding of the received energy. 112
4.6 The ROC of the synchronous single-pulse sensing with LRT detection method. 113
4.7 The ROC of the synchronous multiple-pulse sensing with simple thresholding of the received energy. 114
4.8 The ROC of the synchronous multiple-pulse sensing with LRT detection method. 115
4.9 The ROC of the asynchronous sensing for $p_u = 0.2$ and $p_u = 0.8$ . . . . . 116

4.10 The probability of sensing error versus velocity; $M = 4$ and LRT based synchronous single-pulse sensing is considered. . . . . . . . . . . . . 116
1.1 Background

Globally, the demand for broadband wireless communications is drastically increasing every year. A major factor contributing to this development is the ever-increasing number of users subscribing to broadband internet services using their mobile devices. Moreover, new devices, such as smart-phones and tablets with powerful multimedia capabilities, are entering the market and are creating new demands on broadband wireless access. Finally, new data services and applications are emerging which are key success factors for the mobile broadband experience. All these factors together result in an exponential increase in mobile data traffic in the wireless access system. This trend is expected to continue to a similar extent over the next decade [1].

Current wireless networks are characterized by a static spectrum allocation policy, where governmental agencies assign wireless spectrum to license holders on a long-term basis for large geographical regions. Spectrum is a scarce resource, but measurements reveal that several licensed frequency bands are underutilized most of the time. Recently, because of the aforementioned increase in mobile broadband and spectrum demand, this policy faces spectrum scarcity in particular spectrum bands.
In contrast, a large portion of the assigned spectrum is used sporadically, leading to underutilization of a significant amount of spectrum [2]. Federal Communications Commission (FCC) reports on spectrum utilization reveal that even in the most densely packed urban areas the overall spectrum utilization rarely passes 35% at any one time [3].

The dynamic spectrum access techniques were recently proposed to solve these spectrum inefficiency problems. The key enabling technology of dynamic spectrum access techniques is cognitive radio (CR) technology, which provides the capability to share the wireless channel with licensed users [4]. In this section, we provide a brief background about the cognitive communications and its challenges.

1.1.1 Cognitive Radio Networks

The terms Cognitive Radio and Spectrum Pooling were both defined by Joseph Mitola III within his Dissertation ”Cognitive Radio: An Integrated Agent Architecture for Software Defined Radio” [5]. He suggests the following definition for Cognitive Radio:

”The term cognitive radio identifies the point in which wireless personal digital assistants (PDAs) and the related networks are sufficiently computationally intelligent about radio resources and related computer-to-computer communications to:
(a) detect user communications needs as a function of use context, and (b) to provide radio resources and wireless services most appropriate to those needs.”

Thus a Cognitive Radio is able to automatically select the best and cheapest service for a radio transmission and is even able to delay or bring forward certain
transmissions depending on the currently and supposably soon available resources. In a broad sense, the term cognitive radio refers to various solutions to the problem of spectrum scarcity that seek to overlay, underlay, or interweave the secondary users signals with those of the primary users in such a way that the primary users of the spectrum are as unaffected as possible. In the following, the mechanism of each approach is provided [6].

Interweave Paradigm

The interweave paradigm is based on the idea of opportunistic usage of spectrum bands which was the original motivation for cognitive radio [5, 6]. This paradigm considers the cases where the parts of space–time–frequency are unused by licensed users, referred to as spectrum holes. Therefore, the opportunistic usage of spectrum improves the frequency reuse over the spectrum holes. The interweave technique requires knowledge of the activity information of the noncognitive users known as primary users. Hence, the interweave cognitive radio is an intelligent wireless communication system that periodically monitors the radio spectrum, intelligently detects occupancy in the different parts of the spectrum, and then opportunistically communicates over spectrum holes with minimal interference to the active users [7].

Overlay Paradigm

In this paradigm, the cognitive users can utilize a certain knowledge of primary network and assign part of their power for their own communication. On the other hand, the remainder of the power is used to assist (relay) the primary transmissions [6]. By
careful choice of the power split, the increase in the noncognitive users signal-to-noise power ratio (SNR) due to the assistance from cognitive relaying can be exactly offset by the decrease in the noncognitive users SNR due to the interference caused by the remainder of the cognitive users transmit power used for its own communication [6]. This guarantees that the noncognitive users rate remains unchanged while the cognitive user allocates part of its power for its own transmissions [6].

Underlay Paradigm

The underlay paradigm includes techniques that allow the secondary network to transmit while considering the interference caused by its transmitter to the receivers of all noncognitive users [6]. In this setting, the cognitive radio cannot significantly interfere with the communication of existing (typically licensed) users, who are referred to as primary users. Specifically, the underlay paradigm mandates that concurrent noncognitive and cognitive transmissions may occur only if the interference generated by the cognitive devices at the noncognitive receivers is below some acceptable threshold. While the underlay paradigm is most common in the licensed spectrum (e.g., UWB underlays many licensed spectral bands), it can also be used in unlicensed bands to provide different classes of service to different users [6].

1.2 Research Objective and Outlines

The aim of this research is to employ cognitive communications capability in the primary networks which are enhanced by the emerging techniques in wireless communications. Three recently proposed wireless communication platforms are consid-
ered in this research. These technologies have recently attracted lots of attention in wireless industry to be considered as a part of the next generation wireless standards. Therefore, several mechanisms are proposed in this research to make the utilization of cognitive communications in these networks feasible.

The outline of this dissertation is as follows:

- In Chapter 2, the application of opportunistic usage of cognitive radio in the presence of a $K$-user interference alignment network is provided. The proposed scenario and formulation is firstly presented in Chapter 2 and the feasibility of opportunistic use of spatial holes of interference channel is provided. Then a sensing algorithm is studied to identify the available spatial holes and the numerical simulation of the work is provided in the last section of Chapter 2.

- An application of overlay cognitive radio scheme in a novel technique in wireless communication, spatial modulation, is studied in Chapter 3. The optimal detection and instantaneous error formulation is firstly presented. The aim of Chapter 3 is generally providing a closed-form formulation for the average error probability of the proposed scenario to prove the potential and efficiency of cognitive spatial modulation. These analytical results are also compared with the results from numerical Monte Carlo simulation.

- The application of cognition in devices is studied in another hot topic in the recent wireless communication area, molecular communications, for the first time in Chapter 4. In this chapter, we firstly provide some examples to show the
importance of secondary usage of molecular channels in the future. Then the problem formulation is presented and several sensing mechanisms are proposed for the opportunistic usage of molecular channels. The numerical results are also provided to better show the potential of the cognitive molecular communications.

- The conclusion of results of this work is provided in Chapter 5. Some ideas as the future work of this research are also provided in this chapter for the researcher in this area who are interested in working on this research topic.

In the following, the main objectives of each cognitive radio network is provided briefly. The proposed design and derivations of each work is provided in details in Chapters 2, 3 and 4 of this report.

1.2.1 Opportunistic Usage of Space in $K$-User Interference Network

In Chapter 2, an opportunistic spectrum usage of a primary $K$-user interference network is proposed which is operating under the *interference alignment* (IA) scheme. Unlike conventional CR schemes that search for spectral holes, the proposed CR sensing and communication scheme identifies and utilizes *spatial holes* of the primary interference network. A simple system model is considered which is composed of a point-to-point multi-input multi-output (MIMO) secondary system, while the primary network consists of $K$ pairs of multi-antenna nodes. It is assumed that the primary is able to use the whole achievable *degrees of freedom* (DoFs) with the recently introduced idea of the $K$-user interference alignment [8, 9], in which each
transmitter sends at least one stream of data. The aim is to design a CR scheme
that utilizes the unused DoFs of the primary system for secondary transmission at
no extra cost to the primary. This is done by designing the secondary precoder and
decoder matrices such that no interference is imposed to the primary network, while
the signal-to-interference plus noise ratio (SINR) at the secondary receiver is max-
imized. As the success of this CR scheme depends on successful detection of the
spatial holes in the primary system, a spatial sensing method is derived by taking
advantage of eigenvalue based and generalized likelihood ratio test (GLRT) sensing
schemes in two stages to firstly detect the presence of spatial hole(s) and then to
find the index of inactive primary stream(s). In practice, the proposed CR system
first detects the presence of spatial holes by performing a low complexity eigenvalue
based sensing. If it detects the presence of spatial holes, the CR then proceeds with
finding the indices of the unused DoFs based on the primary channel state information
(CSI), which are then used to design the secondary precoder and decoder matrices.
The secondary precoder and decoder matrices are such that they maximize the sec-
ondary SINR while causing no interference to the primary receivers. Even though the
primary and secondary transmissions are designed for the considered system model,
the proposed sensing scheme is general and applicable to any scenario that involves
spatial sensing. The main contributions of this work can be summarized as:

- the condition on the number of antennas at the secondary link to exploit the
  spatial holes in the primary system without causing interference to the pri-
  mary system as well as the structures of the secondary precoding and decoding
matrices.

- a fast method for detecting spatial holes in the primary system; this method is based on the eigenvalues of the received covariance matrix, and is a coarse sensing scheme that detects the presence of unused or inactive DoFs in the primary system.

- and finally, a fine sensing method that identifies the inactive primary streams; this method is applied after the coarse sensing to accurately determine the index set of inactive data streams in the primary system.

1.2.2 Overlay Cognitive Spatial Modulation

A novel overlay scheme is proposed in Chapter 3 which the CR transmitter utilizes spatial modulation (SM) while assisting the primary network in transmitting its information using amplify-and-forward (AF) relaying. In the proposed scheme, the secondary network transmits its own information in the space domain without causing any interference to the primary network. At the same time, the APM symbols of the primary transmitter is amplified and retransmitted via the active antenna of the secondary transmitter, which will then be detected at the primary receiver. On the other hand, the index of the active transmitting antenna is detected at the secondary receiver. Therefore, the secondary achieves a non-trivial transmission rate without causing any interference to the primary network.

The optimal maximum likelihood (ML) detection is considered at both the primary and secondary receivers to detect the phased-shift keying (PSK) modulated
symbols and the antenna indices, respectively. The instantaneous symbol error rate (SER) in both the primary and secondary receivers is calculated based on summation of pairwise error terms from union bound on the error probability. The average SER (ASER) is then calculated to determine the performance in each domain. Asymptotic analysis is then provided that provides simplified yet insightful equations for the probability of error at high SNRs. Using these asymptotic results, the choice of critical system parameters such as the antenna and APM symbol size is then discussed. The simulation results show that SM can be a viable candidate for overlay CR systems, as it can enhance the performance of the primary system while providing an appealing transmission rate for the secondary network.

1.2.3 Cognitive Molecular Communication

Molecular communication is an emerging communication paradigm for biological nanomachines. The fast grow of molecular communication techniques in variaties of nanomedical applications results in some certain scenarios where there are more than one nanonetworks working in a biological environment. Therefore, the limitation on the type of molecules in carrying information is led to a crunch in molecular channels in the future. This can be considered as an analogy of current spectrum crunch in radio frequency communications which has absorbed huge amount of attentions from wireless communications researchers and regulators. More precisely, the answer to this question can open a new research area in molecular communications; How additional nanonetwork can be involved into a biological system when all transmission
particles are reserved by other nanonetworks?

The work provided in Chapter 4 is the first attempt at employing a nanonetwork with cognitive radio capabilities which is able to communicate with its corresponding users by the aim of molecules which are the same as the molecules for primary transmission. It is assumed that the primary nanonetwork is implemented under concentration-based modulation scheme and the secondary nanonetwork performs the opportunistic approach to sense the availability of primary molecular channel for its own transmission. In the other words, by inspiring from the spectrum sensing in radio frequency communications, the secondary nanonetwork measures the number of molecules as a criteria to decide about the absence or presence of the primary nanonetworks by *molecular energy detection* scheme. Then, when the channel is available, the secondary transmitter sends information such that the interference on the primary receiver is always less than a predefined threshold to prevent the wrong detection at the primary receiver.

We assume that the primary nanonetwork uses the concentration-based modulation scheme and the secondary nanonetwork opportunistically senses the environment for the availability of the primary molecular channel for its own transmission. In the other words, inspired by the spectrum sensing in wireless communications, the secondary nanonetwork measures the number of molecules as a criterion to decide about the presence or absence of the primary nanonetworks using *molecular energy detection*. When no primary transmission is detected, the secondary transmitter sends its information such that the interference at the primary receiver is always below a
predefined threshold thereby preventing possible disruption in the primary operation. In our system model, the energy of the signal at sensing nanodevice is derived based on the Brownian motion of the molecules and the probability of arrival is modeled as an inverse Gaussian random variable. We then divide the sensing mechanism based on the availability of timing information at the sensing nanodevice into two cases: **synchronous** and **asynchronous molecular sensing**. In the synchronous sensing, we provide two mechanisms of sensing by considering the sensing window time smaller or larger than the primary pulse-width, i.e. **single-pulse** and **multiple-pulse sensing**, respectively. We also derive a likelihood ratio test based on the Bayesian approach, when the timing information of the primary is not available. The performance of the proposed sensing schemes in a molecular environment is then evaluated using numerical simulations.

1.3 Literature Survey

This section provides a review of literature in the areas of this research work. Therefore, the publication on the cognitive communications is firstly studied which is the main idea of this dissertation. Then, each of three emerging technologies selected in this research is reviewed individually with or without cognitive capabilities to provide a better background of relevant works in this area.

1.3.1 Cognitive Radio Networks

The current explosion of information and demand for high speed data communication call for novel solutions to utilize the radio resources more efficiently. The cognitive radio (CR) paradigm aims to mitigate this spectrum crunch by opportunistically
exploiting unused licensed spectrum bands that are allocated to the primary communications systems [7]. Since it is a key design objective to not compromise the performance of the primary system, CR schemes are designed to improve spectrum utilization at little or no extra cost to the licensed users. The design of efficient CR schemes also depends on the structure and the architecture of the primary system. For examples, different spectrum sensing and CR communication schemes are necessary when the primary system is a point-to-point network [10], or a cooperative relay network with single [11, 12, 13] or multiple antennas [14].

Any efficient opportunistic CR network must accurately detect the presence or absence of communication in the primary system. Several sensing schemes including energy detection, matched filtering and cyclostationary detection (see e.g. [15] and the references therein) have been proposed in the literature for detecting primary spectral holes. Although the energy detection method does not require a priori information of the primary signal [16], it is not optimal for detecting correlated signals. To overcome the shortcomings of energy detection, eigenvalue based sensing schemes have been suggested [17, 18]. Further, especially for the case of multi-antenna CR nodes, the GLRT has been proposed to utilize the eigenvalues of the sample covariance matrix of the received signal vector without having a priori knowledge of the primary users’ signals [19, 20]. In this method, the primary users’ signals to be detected occupy a subspace of dimension strictly smaller than that of the observation space [19]. In the underlay approach (see e.g. [21, 22, 23]), the secondary transmits in a way that causes small interference to the primary, while in the overlay approach...
(see e.g. \cite{6, 24, 25}), the CR network can also facilitate the primary transmission. In other words, in the overlay approach, which is also considered in this work, the CR assists the primary system by retransmitting its information and thereby improving the performance of the primary system.

1.3.2 Spectrum Sensing in Interference Alignment Networks

The problem of space pooling has been partly studied in \cite{26} and \cite{27} for point-to-point MIMO primary and secondary systems. In these papers, authors present an opportunistic scheme to utilize the unused eigenmodes of the primary channel, and introduce pre- and post-processing schemes. Recently, using the IA techniques in CR networks has been considered \cite{28, 29, 30, 31, 32, 33, 34, 35, 36, 37}. In \cite{28}, authors consider a point-to-point MIMO primary network and propose an iterative algorithm to efficiently design the precoding and decoding matrices for IA in the secondary network. Based on a similar network setup, \cite{29} studies the impact of propagation delay on the DoF of the CR system. In \cite{30}, an outer bound for the total DoF is derived. The achievable DoF when the primary network performs interference suppression is discussed in \cite{31}. With the same network model, authors in \cite{32} provide pre- and post-processing designs of the CR IA network to maintain a target rate for the primary network while maximizing the rate of the secondary link. Recently in \cite{33}, the problem of cooperative spectrum leasing based on a game theoretical approach in an IA network composed of primary and secondary users has been studied. In \cite{34, 35, 36}, authors study the use of IA in underlay CR systems, while the application
of IA in CR femtocell networks has been studied in [37].

1.3.3 Cooperative Spatial Modulation

Spatial modulation (SM) has been recently proposed as a low complexity power-efficient transmission scheme for multi-antenna communications [38, 39]. In this modulation scheme, the modulator uses an amplitude/phase modulation (APM) while employing only one active antenna to convey information in each transmission interval. In other words, SM divides the transmission into two APM and spatial (antenna) domains [40, 41]. A general survey on the design of SM, its advantages and limitations can be found in [42].

The use of space modulation for cooperative relay communications has been recently investigated. For example, [43] analyzes the bit error rate (BER) performance of AF relaying when the source employs space shift keying (SSK) modulation. Similarly, SSK transmission with AF and decode-and-forward (DF) relaying schemes assuming multiple relay nodes has been considered in [44]. In [45], authors propose a cooperative space-time shift keying (STSK) scheme using DF relaying, and the performance of the detector is evaluated. SM using DF relaying with single- and multi-antenna relay nodes is studied in [46] and [47], respectively, and the performance of such systems is derived. In [48], authors provide a DF scheme in which SM is invoked at the source node to divide the bit stream to be transmitted over the APM and antenna domains. Finally, an adaptive SM scheme for spectrum sharing in point-to-point communication systems has been recently proposed in [49] to improve
the performance of the secondary system in terms of energy and spectral efficiency.

The idea of splitting the transmission domains into spatial and APM domains in the second phase of relaying has been recently proposed in [50]. Authors in [50] propose a two phase relaying scheme which consists of multiple distributed single-antenna relays. The source data is retransmitted through spatial domain; similar to distributed space shift keying (DSSK) provided in [51], while the relay node transmits its own data via APM domain. The authors generalize this scheme to distributed spatial modulation (DSM) [50], where the source data must be decoded in each relay independently, and only a particular relay whose identifier (node binary label) coincides with the demodulated symbol is allowed to transmit, while the other relays remain silent. In our proposed scheme, however, the secondary relay node does not require to decode the primary information and the primary APM symbol is only amplified and forwarded to the primary receiver via the activated secondary antenna.

The low complexity of SM scheme is a driving force in employing this scheme in CR networks. For instance, SM has been proposed to be used in the underlay secondary systems [49, 52, 53]. In [49] and [52], authors provide an underlay CR scheme, in which SM is utilized by the secondary system while the interference to a single-antenna primary node is controlled by the secondary transmitter. By considering a feedback scheme from the primary to the secondary, the secondary is able to keep its interference to the primary below a threshold by adaptively employing a certain set of its antennas. The so-called quadrature SM scheme is considered in [53], in which the interference at the multiple primary users is controlled by transmit power of the
active antenna at the secondary transmitter. A similar underlay scheme in which
the secondary transmitter uses SSK is proposed in [54], and the error performance
is evaluated when the interference to multiple primary users is kept below a thresh-
old. Recently, some papers have considered the use of the SM in overlay CR systems
[55, 56]. In [55], the authors consider DF relaying at the secondary transmitter and
the performance of the system in terms of the outage probability is studied. The DF
relaying scheme is also considered in [56] and the error performances of the primary
and secondary systems are evaluated numerically.

1.3.4 Cognitive Molecular Communications

Molecular communication has recently gained attention as a mechanism to transmit
information in nanoscale systems, especially in scenarios where conventional electro-
magnetic communication is not feasible or safe [57, 58, 59, 60]. Inspired by biologi-
cal exchange of information in nature, this communication technique uses carrier
molecules or lipid vesicles released into an aqueous or gaseous medium to convey
information from a nanoscale transmitter to a receiver. The application of molecu-
lar communications is growing fast, especially in interconnection between nanoscale
devices, such as intra-body sensing and actuation networks, and environmental con-
trol of toxic agents and pollution [61, 62, 63]. The information can be encoded into
the carrier molecules in at least three different ways: concentration of molecules,
type of molecules and time of release of molecules. By selecting one or a combina-
tion of these parameters, different modulation schemes have been recently proposed

16
In recent years, the progress in nanotechnology has established the foundations for implementing nanomachines capable of carrying out simple but significant tasks \[63, 69\]. It is expected that in some scenarios, several nanonetworks should coexist in the same environment. Choosing the proper type of molecule for data transmission in a molecular communication system is a challenging issue. The choice should not be harmful to the environment and, at the same time, should provide a reliable transmission without causing disruption in the normal operation of other nanonetworks in the environment. For instance, the choice of information carrying particles for in-body molecular communications is very critical and the choices of molecule type is limited due to the complexity of the biologicals cells and organisms \[70, 66\]. Authors in \[64\] propose the molecular shift keying (MoSK) scheme based on hydrofluorocarbon as information carrying particle; while \[65\] suggests aldohexose isomers considering the effect of harmful isomers of hydrofourocarbons on the human body. While several promising applications of molecular communications for in-body information exchange, e.g. drug delivery, are envisioned, the limitation on the types of proper molecules for data transmission can be a limiting factor \[63\].

To our knowledge, this work is the first attempt at introducing the concept of cognitive molecular communications, in which a secondary nanonetwork with cognitive capabilities transmits its own information using the same type of carrier molecule as that of a primary nanonetwork. This is achieved without any disruption in the normal operation of the primary system.
CHAPTER II

COGNITIVE RADIO COMMUNICATION IN THE PRESENCE OF A K-USER INTERFERENCE PRIMARY NETWORK

In this chapter, we propose opportunistic spectrum usage of a primary $K$-user interference network operating under the *interference alignment* (IA) scheme. Our proposed CR sensing and communication scheme identifies and utilizes *spatial holes* of the primary interference network. The aim is to design a CR scheme that utilizes the unused DoFs of the primary system for secondary transmission at no extra cost to the primary. So, we firstly introduce the system model and problem formulation. Then, the precoder and decoder matrices are provided such that SINR at the secondary receiver is maximized. Then fast and fine spatial hole sensing schemes are provided while the last section of this chapter provides the numerical simulation of the proposed method.

2.1 System Model

In this section, we firstly provide the proposed cognitive radio model and the parameters of the network model are defined. Then, the primary interference alignment network is described and the secondary system formulation is provided to be used in the next section.
2.1.1 Overview of the Scenario

We consider a $K$-user MIMO interference channel shown in Fig. 2.1 as the primary system, where the IA is done at the primary transmitters having full CSI [71]. In this network, each transmitter transmits one or a few streams of data to its corresponding receiver. However, we assume that during some transmission intervals one or some of the primary transmitters does not have any data to transmit, and therefore, some DoFs of the primary system remain unused at those time intervals. We consider a pair of multi-antenna secondary nodes to opportunistically utilize the unused DoFs without causing interference to the primary system.

It is assumed that the secondary system has full CSI of all the primary links, primary to secondary, secondary to primary, as well as the link between the secondary
transmitter and receiver. Although acquiring full CSI at the secondary system is
difficult in practice, there are a number of ways to make this possible. Based on
the channel reciprocity, CR nodes are able to estimate the channel information from
the primary to secondary nodes and vice versa using pilot signals of the primary
network. A critical assumption in the realization of a distributed IA network is the
availability of robust feedback channels [72]. In addition, in the proposed scheme,
the CR network requires the knowledge of precoder and decoder matrices of the
primary network. A practical way to obtain this information in the CR network
using cognitive pilot channel (CPC) has been introduced [73], and currently standard
organizations are considering using such channels to broadcast useful information of
primary to available secondary networks for opportunistic or underlay utilization of
the spectrum [74].

2.1.2 Primary IA network

We assume that in the primary network, the \(k^{th}\) transmitter and receiver pair have
\(M^{[k]}\) and \(N^{[k]}\) antennas, respectively. The DoF for the signal of the \(k^{th}\) pair is defined
by \(d^{[k]} \leq \min(M^{[k]}, N^{[k]}), \ k \in \mathcal{K}\) where \(\mathcal{K} \triangleq \{1, 2, ..., K\}\) is the set of all primary pairs
[8, 9]. The receive signal at the \(k^{th}\) primary receiver can be written as:

\[
Y^{[k]} = \sum_{l=1}^{K} H^{[kl]} X^{[l]} + Z^{[k]}, \quad \forall k \in \mathcal{K} \tag{2.1}
\]

where \(X^{[l]}\) is the \(M^{[l]} \times 1\) transmit signal vector of the \(l^{th}\) primary transmitter, \(Y^{[k]}\)
is the \(N^{[k]} \times 1\) receive vector, and \(Z^{[k]}\) is the complex additive white Gaussian noise
(AWGN) vector at the \(k^{th}\) receiver with independent and identically distributed (i.i.d)
circularly symmetric complex Gaussian entries with zero mean and variance $\sigma_z^2$. $\mathbf{H}^{[kl]}$ is the $N[k] \times M[l]$ matrix of channel coefficients between the $l^{th}$ transmitter and the $k^{th}$ receiver as shown in Fig. 2.1. For simplicity, it is assumed that all channel matrices are full rank; however, the proposed scheme can be easily extended to the case of rank deficient matrices.

Precoder at primary

The signal vector transmitted by the $k^{th}$ transmitter can be written as:

$$
\mathbf{X}[k] = \sum_{d=1}^{d[k]} \mathbf{V}[k] \overline{X}_d = \mathbf{V}[k] \overline{X}[k],
$$

(2.2)

where $\overline{X}[k]$ is a $d[k] \times 1$ vector, $\overline{X}_d$ is the $d^{th}$ entry of vector $\overline{X}[k]$ and $\mathbf{V}[k]$ is an $M[k] \times d[k]$ precoder matrix whose columns are the orthogonal basis for the transmitted signal space of the $k^{th}$ transmitter [9]. We assume that the average transmit power at the $k^{th}$ transmitter is $\mathbb{E}[||X[k]||^2] = p[k]$ and that the power is allocated uniformly to all data streams.

Decoder at primary

The interference suppression at the primary receiver can be utilized to construct an $N[k] \times d[k]$ decoder matrix $\mathbf{U}[k]$ with orthogonal columns to minimize the interference in the desired signal subspace at the $k^{th}$ receiver. Therefore, the receive signal after decoder for the $k^{th}$ receiver is:

$$
\mathbf{Y}^{[k]} = \mathbf{U}^{[k] \dagger} \mathbf{Y}^{[k]}.
$$

(2.3)
Feasibility of alignment

If the interference is aligned into the null space of $U^{[k]}$, then the following condition must be satisfied:

$$U^{[k]}H^{[kj]}V^{[j]} = 0_{d^{[k]} \times d^{[j]}}, \forall j \neq k, \tag{2.4}$$

$$\text{rank}(U^{[k]}H^{[kk]}V^{[k]}) = d^{[k]}, \forall k \in \mathcal{K}, \tag{2.5}$$

while columns of the precoding and decoding matrices construct orthogonal basis sets; i.e.:

$$V^{[k]} : M^{[k]} \times d^{[k]}, V^{[k]}V^{[k]^\dagger} = I_{d^{[k]}}, \tag{2.6}$$

$$U^{[k]} : N^{[k]} \times d^{[k]}, U^{[k]^\dagger}U^{[k]} = I_{d^{[k]}}, \tag{2.7}$$

Numerical solutions

Although closed-form solutions have been found for the IA problem in three-user interference channels [9], the closed-form solution for the general case of $K$-user interference channel is unknown and such a problem is NP-hard. There are, however, some numerical iterative algorithms suggested in the literature to design the precoder and decoder in the case of $K$-user IA with full CSI [75, 76, 77]. In this work, and in the numerical results section, we consider the method provided in [75] to derive the precoding and decoding matrices in the primary network. This method considers the reciprocity of the channels and assigns the eigenvector corresponding to the the smallest eigenvalue of the receive matrix as precoder. In the next step, the decoder at the receiver can be considered as the precoder of the reciprocal channel. This iteration continues until the solution converges to an IA solution.
2.1.3 Secondary MIMO Link

We consider a pair of multi-antenna transmitter and receiver as the secondary network (the pair with index $k = 0$ shown in Fig. 2.1). It is assumed that the secondary network works at the same frequency and time as that of the primary network. The $M[0] \times d[0]$ matrix $V[0]$ and the $N[0] \times d[0]$ matrix $U[0]$ are the precoding and decoding matrices of the secondary network, respectively. The number of streams that can be transmitted by the CR system is $d[0]$. Our aim is to find these two matrices to minimize the interference imposed to the primary network while maximizing the transmission rate of the secondary system.

Interference leakage

We assume that the interferences in the primary network have been aligned and that the $k^{th}$ primary transmitter has $d[k]$ degrees of freedom. The interference leakage at the $k^{th}$ primary receiver due to the secondary transmission is the summation of all interferences from secondary streams to the $k^{th}$ primary receiver; i.e.:

$$I_{CR}^{[k]} = \text{Tr} \left[ \frac{p[0]}{d[0]} U[k]^\dagger H[k0] V[0] V[0]^\dagger H[k0]^\dagger U[k] \right],$$

(2.8)

where it is assumed that the total secondary transmit power, $p[0]$, is allocated uniformly to all data streams.
SINR at CR receiver

The SINR of the $l^{th}$ secondary data stream at the secondary receiver can be written as [75]:

$$\gamma_{l}^{CR} = \frac{p[0] U_{||l}^{[0]} H_{\parallel l}^{[0]} V_{\parallel l}^{[0]} V_{\parallel l}^{[0] \dagger} H_{\parallel l}^{[0]} U_{\parallel l}^{[0] \dagger}}{d[0] B_{l} U_{\parallel l}^{[0] \dagger} U_{\parallel l}^{[0]} U_{\parallel l}^{[0] \dagger} H_{\parallel l}^{[0]} V_{\parallel l}^{[0] \dagger} H_{\parallel l}^{[0]} V_{\parallel l}^{[0]} \sigma_{z}^{2} I_{N[0]}},$$

where

$$B_{l} = \sum_{j=0}^{K-1} \frac{d[j]}{d[l]} \sum_{d=1}^{d[j]} H_{\parallel d}^{[0]} V_{\parallel d}^{[0]} V_{\parallel d}^{[0] \dagger} H_{\parallel d}^{[0]} \dagger, \quad (2.10)$$

where we have not considered the possibility of successive interference cancellation in the SINR equation.

2.2 Secondary Interference Alignment

In this section, we consider the design of the secondary precoder and decoder matrices in the presence of the primary $K$-user interference network. We, first, derive the sufficient condition on the number of antennas at the secondary transmitter in order to utilize the spatial holes for its own transmission without causing interference to primary. Then, we provide the pre- and post-processing (precoding and decoding) matrices to maximize the SINR at the CR receiver while the interference leakage to the primary is zero.

2.2.1 Precoder at Secondary Transmitter

We consider the summation of interference terms from the secondary transmitter to the primary receivers. Our aim is to design the precoder at the secondary transmitter
such that the interference leakage to the primary receivers is minimized or optimally nullified. Therefore, the optimization problem to find the secondary precoder can be written as:

$$\min_{V_0: M_0 \times d_0} \sum_{k=1}^{K} I_{CR}^{[k]}$$

s.t. $V_0^\dagger V_0 = I_{d_0}$,

where the precoder matrix $V_0$ at the CR transmitter is assumed to be a unitary matrix to preserve the secondary transmit power constraint. While such a choice will minimize the interference, it does not always guarantee zero interference at each primary receiver. Instead, we shall first state the sufficient condition on the number of transmit antennas at the secondary to completely nullify the interference leakage to the primary receivers, and then propose an alternative precoder design to achieve zero interference.

**Lemma 1 (Sufficient condition for zero interference).** The number of secondary transmit antennas is limited by

$$M_0 \geq d_0^0 + \sum_{i=1}^{K} d_i^0$$

where $d_i^0$ is the number of active streams of the $i^{th}$ primary transmitter.

**Proof.** To realize the interference-free channels from the secondary transmitter to the primary receivers, one can design the secondary precoder matrix to align the
Figure 2.2: A secondary system in the presence of a primary 3-user IA system; left: all the primary transmitters are active and there is no unused DoFs. right: the second primary transmitter is off and there is one unused DoF that can be used by the secondary system.

The size of the $i^{th}$ matrix in $P$ is $d[i] \times M[0]$, and therefore, when all streams of the primary transmitters are active, the rank of $P$ can be written as:

$$\text{rank}(P) = \min(K \sum_{i=1}^{K} d[i], M[0]).$$

(2.14)
By denoting the number of active streams in the $i^{th}$ primary transmitter as $d^{[i]}_A$, the rank of $P$ becomes:

$$\text{rank}(P) = \min(K \sum_{i=1}^{K} d^{[i]}_A, M^{[0]}), \quad (2.15)$$

and if $\sum_{i=1}^{K} d^{[i]}_A < M^{[0]}$, then $\min(K \sum_{i=1}^{K} d^{[i]}_A, M^{[0]}) = \sum_{i=1}^{K} d^{[i]}_A$. So, one can simply find the dimension of the null space of $P$ as:

$$\text{null } (P) = M^{[0]} - \sum_{i=1}^{K} d^{[i]}_A. \quad (2.16)$$

Therefore, the number of transmit antennas and data streams at the secondary should satisfy the following equation:

$$\text{null } (P) = M^{[0]} - \sum_{i=1}^{K} d^{[i]}_A \geq d^{[0]}, \quad (2.17)$$

To have zero interference at all the primary receivers, we should then have (2.12). \hfill \Box

**Remark 1.** The inequality (2.12) of Lemma 1 can be considered as the maximum number of streams that the secondary system is able to transmit which is limited by $d^{[0]} \leq M^{[0]} - \sum_{i=1}^{K} d^{[i]}_A$. More precisely, if we assume the same number of secondary transmit antennas as the DoFs of the primary system, i.e., $\sum_{i=1}^{K} d^{[i]}$, the CR network is able to use all the unused DoFs of the primary.

The $c^{th}$ column of the precoder matrix $V^{[0]}$ ($c = 1, \ldots, d^{[0]}$) can be defined by the eigenvector corresponding to one of the $\sum_{i=1}^{K} \left( d^{[i]} - d^{[i]}_A \right)$ zero eigenvalues of $P$; i.e.:

$$V^{[0]}_c = E_{d}[P] \quad (2.18)$$

where $l = \sum_{i=1}^{K} d^{[i]}_A + c$.  

27
Example: As shown in Fig. 2.2, consider a 3-user MIMO IA primary system in which all the nodes have two antennas, and each primary transmitter sends one stream of data to its corresponding receiver. We have also assumed a $3 \times 3$ secondary link. If all the primary transmitters are active, there will be no unused DoFs to be used by the secondary. However, when only two or less primary transmitters are active, knowing that the number of transmit antennas at the secondary satisfy the condition in (2.12), the CR network is able to use the unused DoF(s) of the primary system. For the case of one unused DoF, the secondary transmitter can select the subspace corresponding to the null space of the secondary to primary interference channel, i.e., $V^{[0]} = E_3[P]$ as its precoder.

Remark 2 (Underlay CR Approach). If the null space of $P$ is empty, no secondary precoder exists that nullifies the interference at all the primary receivers. In such a case, one can consider an underlay CR approach in which the interference to all the primary receivers can be kept below a predefined threshold $\eta$; i.e.:

$$I^{[1]}_{\text{CR}} \simeq I^{[2]}_{\text{CR}} \simeq ... \simeq I^{[K]}_{\text{CR}} \leq \eta. \quad (2.19)$$

In this work, we do not consider such an underlay approach, and assume that there are some unused DoFs in the primary system that can be utilized by the secondary. However, without considering the individual constraint on the interference leakage in (2.19), the interference minimization problem (2.11) can still be considered while the number of transmit antennas of the secondary network does not satisfy the condition in Lemma 1. In this case, considering $\text{Tr}[A + B] = \text{Tr}[A] + \text{Tr}[B]$ and $\text{Tr}[AB] =
Tr[BA], we can rewrite the objective function of (2.11) as:

\[
\min_{\hat{V}^0} \text{Tr} \left[ \hat{V}^{0\dagger} Q \hat{V}^0 \right] \tag{2.20}
\]

s.t. \( \hat{V}^{0\dagger} \hat{V}^0 = I_{d^{[0]}} \),

where \( Q \) is an \( M^{[0]} \times M^{[0]} \) matrix such that:

\[
Q = \sum_{k=1}^K \frac{p^{[0]}_{\check{d}^{[0]}}}{d^{[0]}} H^{[k\check{0}]} U^{[k]} U^{[k\check{0}]} \dagger H^{[k\check{0}]} \tag{2.21}
\]

Since \( Q \) is Hermitian, by using trace minimization [78, p.191], the total interference leakage is the summation of \( d^{[0]} \) smallest eigenvalues of \( Q \).

2.2.2 Decoder at Secondary Receiver

The principal goal of the CR network is to agilely utilize the unused DoFs in such a way to optimize a performance objective. In this work, we consider the secondary SINR maximization approach to design the secondary decoder.

**Lemma 2** (Decoder matrix). The \( l \)th column of the decoder matrix \( U^{[0]} \) to maximize the SINR can be derived as:

\[
U^{[0]}_{l|l} = \frac{(B_l)^{-1} H^{[00]} \hat{V}^0_{l|l}}{\| (B_l)^{-1} H^{[00]} \hat{V}^{0}_{l|l} \|}, \tag{2.22}
\]

and the maximum SINR achieved by this solution is:

\[
\gamma^\text{CRmax}_l = \frac{p^{[0]}_{\check{d}^{[0]}}}{d^{[0]}} \hat{V}^{0\dagger}_{l|l} H^{[00]} \dagger (B_l)^{-1} H^{[00]} \hat{V}^{0}_{l|l}. \tag{2.23}
\]

**Proof.** The proof is in [75, Section V.C].
Remark 3. If we define the normalized vector $H_l = H_{[0]}^{[0]} V_{[0]}^{[0]}$, the instantaneous SINR at the CR receiver can be bounded by the Kantorovich matrix inequality [79]:

$$
(H_l^H B_l H_l)^{-1} \leq \gamma^{\text{CRmax}} \leq \frac{(\lambda_1^l + \lambda_{N[0]}^l)^2}{4 \lambda_1^l \lambda_{N[0]}^l} \left( H_l^H B_l H_l \right)^{-1},
$$

(2.24)

where $\lambda_1^l$ and $\lambda_{N[0]}^l$ are the largest and the smallest eigenvalues of $B_l$, respectively.

Lemma 3 (Number of receive antennas). Assuming single stream transmission in the CR network, the average SINR at the CR receiver is an increasing function of the number of receive antennas at the secondary.

Proof. To see the effect of the number of secondary receive antennas, $N[0]$, on the average SINR, we show that by adding one additional antenna at the secondary receiver, the average SINR increases. We first define the new vector $\tilde{H}_l$ and matrix $\tilde{B}_l$ of the receiver with one additional antenna as:

$$
\tilde{H}_l = \begin{bmatrix} H_l \\ h_l \end{bmatrix}, \quad \tilde{B}_l = \begin{bmatrix} B_l & B_1 \\ B_1^\dagger & b_2 \end{bmatrix},
$$

(2.25)

where $\tilde{H}_l$ and $H_l$ are vectors of size $(N[0] + 1) \times 1$ and $N[0] \times 1$, respectively, and matrices $\tilde{B}_l$ and $B_l$ are of size $(N[0] + 1) \times (N[0] + 1)$ and $N[0] \times N[0]$, respectively. Hence, the inverse of partitioned matrix $\tilde{B}_l$ can be written as:

$$
\tilde{B}_l^{-1} = \begin{bmatrix} \hat{B}^{-1} & \hat{B}_1 \\ \hat{B}_1^\dagger & \hat{b}_2 \end{bmatrix},
$$

(2.26)
where \( \hat{\mathbf{B}} = \mathbf{B}_t - \frac{B_1 B_1^\dagger}{b_2} \), and \( \hat{B}_1 \) and \( \hat{b}_2 \) are well-defined vector and scalar, respectively [78]. The average SINR at the secondary receiver in (2.23) can be written as:

\[
E[\tilde{\gamma}_t^{\text{CRmax}}]_{(N^0)+1} = \frac{p^0}{d^0} E[\hat{H}_t \hat{\mathbf{B}}_1^{-1} \hat{H}_t] = \frac{p^0}{d^0} \left( E[H_t \hat{\mathbf{B}}_1^{-1} H_t] + 2\Re\{E[h_t^* \hat{B}_1^\dagger H_t]\} + E[|h_t|^2 \hat{b}_2]\right). \tag{2.27}
\]

Considering single stream transmission by the secondary, and based on the definition of interference matrix \( \mathbf{B}_t \) in (10), there is no common random variable in \( \mathbf{B}_t \) and \( H_t \). Therefore, \( h_t, H_t \) and \( \hat{B}_1 \) are independent, and the second term in (A.3) is zero because \( E[h_t] = 0 \). Furthermore, the third term in (A.3) is always positive. To prove the lemma, it suffices to compare the first term in this equation with the average SINR in the case of \( N^0 \) number of antennas.

For this purpose, we first prove that \( \hat{\mathbf{B}} \) is positive definite (PD); i.e. for all non-zero vectors \( X, X^\dagger \hat{\mathbf{B}} X > 0 \). Considering that matrix \( \hat{\mathbf{B}}_t \) is PD, for any given non-zero vector \( X^H = \begin{bmatrix} X_1^H & x_2^* \end{bmatrix} \):

\[
X^\dagger \hat{\mathbf{B}}_t X = X_1^\dagger \left( b_2 \mathbf{B}_t - B_1 B_1^\dagger \right) X_1 + ||B_1^\dagger X_1 + b_2 x_2||^2. \tag{2.28}
\]

By choosing \( x_2 = -B_1^\dagger X_1 / b_2 \), it follows that the first term in (A.4), i.e. \( X_1^\dagger \left( b_2 \mathbf{B}_t - B_1 B_1^\dagger \right) X_1 \), is always positive. Therefore, the matrix \( \left( b_2 \mathbf{B}_t - B_1 B_1^\dagger \right) \) is PD and hence, \( \hat{\mathbf{B}} \) is PD.

Considering the fact that \( \mathbf{B}_t \succeq \hat{\mathbf{B}} \) and since both matrices are PD, we can conclude that \( \mathbf{B}_t^{-1} \preceq \hat{\mathbf{B}}^{-1} \) [32]. The first term in (A.3) is always larger than the average SINR when the secondary receiver has \( N^0 \) antennas. So, \( E[\tilde{\gamma}_t^{\text{CRmax}}]_{(N^0)+1} \geq E[\tilde{\gamma}_t^{\text{CRmax}}]_{(N^0)} \), and the average SINR is an increasing function of the number of antennas at the secondary receiver. \( \square \)
Even though Lemma 3 concerns with the case of single data stream at the secondary network, its statement is valid for the case of transmission of more than one stream. We have shown this with numerical simulations in Chapter 4.

Example: As depicted in Fig. 2.2, the received vectors at the secondary receiver from two active primary transmitters only span two dimensions out of three available dimensions of receiver. At high SNRs, the optimal decoder is the vector perpendicular to the two-dimensional plane which is spanned by two interference vectors from active primary transmitters (see equation (2.22)). By increasing the number of antennas of the secondary receiver, there will be more degrees of freedom to choose the optimal decoder. This helps in increasing the received SINR of the secondary as stated in Lemma 3.

2.3 Null Space Sensing and Detection

In the previous section, we showed that the unused DoFs of the primary network can be utilized by the secondary MIMO system. In this section, we provide a fast sensing technique to identify the availability of null space of the primary. In the proposed model, the CR receiver has the role of a fusion center and has to detect the availability of spatial holes (null spaces) of the concurrent primary $K$-user IA system. In this scheme, a spectrum sensing method at the CR fusion center determines the availability of the spatial holes. Note that using this scheme, the fusion center cannot identify the indices of the unused DoFs, or equivalently, inactive data streams. The main reason for applying this method is to decrease the computational load in fusion
center to sense the availability of inactive DoFs. To find the indices of the inactive primary streams, a search method will be introduced in the next section.

The null space sensing problem can be formulated as a hypothesis test. In the proposed hypothesis test, \( \mathcal{H}_0 \) implies that there exists an inactive data stream and \( \mathcal{H}_1 \) indicates that the primary null space is empty. Considering the set of received vectors from the primary transmitters, we can rewrite the hypotheses as:

\[
\mathcal{H}_0 : Y[n] = \mathcal{X}_0[n] + Z[n], \quad n = 0, ..., L - 1, \tag{2.29}
\]
\[
\mathcal{H}_1 : Y[n] = \mathcal{X}_1[n] + Z[n], \quad n = 0, ..., L - 1,
\]

where:

\[
\mathcal{X}_0[n] = \sum_{k=1}^{K} \sum_{j=1}^{d_k} \sum_{j \neq l[i]} \mathbf{H}_{j}^{[k]} \mathbf{V}_{j}^{[k]} \bar{\mathbf{X}}_{j}^{[k]}[n], \tag{2.30}
\]

when the \( l^{th} \) stream of the \( i^{th} \) user is inactive, and:

\[
\mathcal{X}_1[n] = \sum_{k=1}^{K} \sum_{j=1}^{d_k} \mathbf{H}_{j}^{[k]} \mathbf{V}_{j}^{[k]} \bar{\mathbf{X}}_{j}^{[k]}[n], \tag{2.31}
\]

in the case that all the streams are active. \( Z[n] \) is the zero mean Gaussian noise vector at the fusion center. For estimating the covariance matrix of the received signal, we firstly define the smoothing factor \( T \) such that \( T \geq N[0] \) [18]. By considering the received and noise vectors over \( T \) consecutive sample times, we can define the
following matrices:

\[
\bar{Y}[n] \overset{\text{def}}{=} [Y^T[n], Y^T[n-1], \ldots, Y^T[n-T+1]]^T,
\]

\[
\bar{X}_i[n] \overset{\text{def}}{=} [X_i^T[n], X_i^T[n-1], \ldots, X_i^T[n-T+1]]^T, i = 0, 1,
\]

\[
\bar{Z}[n] \overset{\text{def}}{=} [Z^T[n], Z^T[n-1], \ldots, Z^T[n-T+1]]^T.
\]

Assuming that there are a total of \( L \) such matrices, the covariance matrices normalized by the known noise variance for the \( L \) collected sampled matrices can be written as:

\[
R_Y = \frac{1}{\sigma_z^2} \mathbb{E} \{ \bar{Y}[n] \bar{Y}^\dagger[n] \} \approx \frac{1}{L \sigma_z^2} \sum_{n=T-1}^{T+L-2} \bar{Y}[n] \bar{Y}^\dagger[n], \quad (2.33)
\]

\[
R_{X_i} = \frac{1}{\sigma_z^2} \mathbb{E} \{ \bar{X}_i[n] \bar{X}_i^\dagger[n] \} \approx \frac{1}{L \sigma_z^2} \sum_{n=T-1}^{T+L-2} \bar{X}_i[n] \bar{X}_i^\dagger[n], \quad (2.34)
\]

\[
R_Z = \frac{1}{\sigma_z^2} \mathbb{E} \{ \bar{Z}[n] \bar{Z}^\dagger[n] \} \approx \frac{1}{L \sigma_z^2} \sum_{n=T-1}^{T+L-2} \bar{Z}[n] \bar{Z}^\dagger[n]. \quad (2.35)
\]

Since the noise and data are independent, the received covariance matrix in hypothesis \( \mathcal{H}_0 \) can be written as:

\[
R_Y = R_{X_0} + R_Z. \quad (2.36)
\]

Therefore, the received covariance matrix is the summation of two covariance matrices. If we assume that the secondary has the same number of antennas as the total DoFs of the primary network, it can be simply shown that the \( N^{[0]} \)th (smallest) eigenvalue of \( R_{X_0} \) is zero. We can employ the eigenvalue analysis methods for sensing the availability of null space. Using properties of eigenvalues for spectrum sensing has been studied for other networks [80, 81, 82]. Here, we propose an eigenvalue based sensing method for a \( K \)-user IA system as a novel application of this sensing method.
Lemma 4 (Fast Eigenvalue Sensing). The probability of a false alarm (PFA) for a predefined threshold $\eta$ is bounded as:

$$\Pr(\lambda_{\min}(R_Z) > \eta) \leq P_{FA} \leq \Pr(\lambda_{\max}(R_Z) > \eta),$$

where $P_{FA} = \Pr(\lambda_{\min}(R_Y) > \eta|\mathcal{H}_0)$ and $N[0] = \sum_{i=1}^{K} d[i]$.

Proof. The Weyl’s inequality for the summation of eigenvalues in hypothesis $\mathcal{H}_0$ states that:

$$\lambda_{\min}(R_{X_0}) + \lambda_{\min}(R_Z) \leq \lambda_{\min}(R_Y) \leq \lambda_{\min}(R_{X_0}) + \lambda_{\max}(R_Z)$$

(2.38)

If we assume that the number of antennas at the fusion center is equal to the total number of data streams of the $K$-user IA system ($N[0] = \sum_{i=1}^{K} d[i]$), the smallest eigenvalue of the covariance matrix, $R_{X_0}$, shows the presence or absence of a spatial hole. Then, the inequality (2.38) can be changed to:

$$\lambda_{\min}(R_Z) \leq \lambda_{\min}(R_Y) \leq \lambda_{\max}(R_Z)$$

(2.39)

Note that $R_Z$ is statistically known at CR fusion center. In such a case, the PFA for the a predefined threshold $\eta$ is bounded as:

$$\Pr(\lambda_{\min}(R_Z) > \eta) \leq P_{FA} \leq \Pr(\lambda_{\max}(R_Z) > \eta)$$

(2.40)

where $P_{FA} = \Pr(\lambda_{\min}(R_Y) > \eta|\mathcal{H}_0)$. \hfill \Box

The noise vectors are $N[0]$-variate complex Gaussian vectors with covariance matrix $\Sigma = \sigma_z^2 I$ and are assumed to be independent. Therefore, the covariance
matrix $R_Z$, as calculated in (2.35), has a central complex Wishart distribution with $LT$ degrees of freedom and covariance matrix $\Sigma$, i.e., $R_Z \sim \mathcal{W}_{N[0]}(LT, \Sigma)$ [83]. Note that the distribution of random matrix $R_Y$ is not available. However, according to (2.37), the threshold $\eta$ can be set properly by knowing the distribution of the smallest and largest eigenvalues of $R_Z$. Since $R_Z$ has a Wishart distribution, its covariance matrix $\Sigma$ is a sufficient statistic for setting the threshold $\eta$ [84, Chap. 7].

As explained, the proposed sensing method only detects the presence or absence of unused DoFs. Therefore, this scheme only serves a coarse detection step. The detection of the exact number of inactive data streams and their indices is provided in the next section. Note that if the coarse sensing step cannot make the correct decision about the presence of spatial hole, the fine detection in next step will not be performed which translates in a performance degradation of the total sensing scheme. Hence, the PFA plays a crucial role in the sensing. We, next, consider the upper bound on the PFA to find a suitable threshold $\eta$.

The cumulative distribution function (CDF) of the largest or smallest eigenvalue of a Wishart matrix is a well-known problem in random matrix theory [83]. For the case of central Wishart matrices, Khatri’s result provides the closed-form function as [85, 86]:

$$\Pr(\lambda_{max}(R_Z) \leq \eta) = \frac{|\Psi(\eta)|}{\prod_{k=1}^{N[0]} \Gamma(LT - k + 1) (N[0] - k + 1)}$$

(2.41)

where $|\cdot|$ denotes the determinant, and $\Psi(\eta)$ is an $N[0] \times N[0]$ Henkel matrix function.
of \( \eta \in (0, \infty) \) with entries given by

\[
\{ \Psi (\eta) \}_{i,j} = \Gamma \left( LT - N_0 + i + j - 1, L\eta \right), i = 1, \ldots, N_0 \tag{2.42}
\]

where \( \Gamma(\cdot, \cdot) \) is the incomplete Gamma function. While Khatri’s formulation provides the closed form CDF, it is difficult to use it to derive a closed-form easy-to-use equation for \( \eta \). Therefore, an approximation method for deriving the distribution of the largest and smallest eigenvalues has been recently proposed in literature [87]. When \( \lim_{LT \to \infty} \frac{N_0}{LT} = y \) where \( 0 < y < 1 \), the CDFs of the largest and smallest eigenvalues approximately converge to the Tracy-Widom distribution of order two; i.e. for the largest eigenvalue:

\[
\Pr(\lambda_{\text{max}}(R_Z) \leq \eta) \approx F_2 \left( \frac{L\eta - \left( \sqrt{LT} + \sqrt{N_0} \right)^2}{\left( \sqrt{LT} + \sqrt{N_0} \right) \left( \sqrt{\frac{1}{LT}} + \sqrt{\frac{1}{N_0}} \right)^{\frac{1}{3}}} \right) \tag{2.43}
\]

and for the smallest eigenvalue:

\[
\Pr(\lambda_{\text{min}}(R_Z) \leq \eta) \approx F_2 \left( \frac{L\eta - \left( \sqrt{LT} - \sqrt{N_0} \right)^2}{\left( \sqrt{LT} - \sqrt{N_0} \right) \left( \sqrt{\frac{1}{LT}} - \sqrt{\frac{1}{N_0}} \right)^{\frac{1}{3}}} \right) \tag{2.44}
\]

where \( F_2 \) is the Tracy-Widom distribution of order two and can be written as:

\[
F_2 (\lambda) = \exp \left( - \int_{\lambda}^{\infty} (x - \lambda) q^2(x) dx \right), \tag{2.45}
\]

where \( q(x) \) is the solution to the non-linear Painlevé’s equation of type II, i.e.:

\[
q''(x) = xq(x) + 2q^3(x). \tag{2.46}
\]
Using this approximate method, and assuming the availability of the inverse of Tracy-Widom distribution function, the threshold value can be derived as:

\[ \eta = \frac{1}{L} \left( \sqrt{LT} + \sqrt{N[0]} \right) \left( \sqrt{\frac{1}{LT}} + \sqrt{\frac{1}{N[0]}} \right)^{\frac{3}{2}} F_2^{-1}(1 - P_{FA}) + \frac{1}{L} \left( \sqrt{LT} + \sqrt{N[0]} \right)^2. \]  

(2.47)

Thus, if we consider the minimum eigenvalue of the received covariance matrix in (2.38) as the test statistic, since the degree of freedom of the Wishart distribution in the received covariance matrix is the multiplication of two sampling parameters, \(T\) and \(L\), the PFA has sharp decreasing behaviour. By proper selection of the threshold value based on a reasonable PFA, this method provides a fast scanning scheme while keeping the probability of correct detection of a spatial hole high.

It is worth mentioning that, for the sake of simplicity, we considered the hypothesis \(H_0\) in (2.30) as the case when there is only one inactive primary stream. The proposed sensing method, however, is general and includes the case of the detection of more than one spatial hole. In this case, the minimum eigenvalue of covariance matrix of the signal, \(R_{x_0}\), in (2.38) is still zero. We will show in the simulation results that in this case the PFA can be slightly lower than the case of only one inactive stream.

2.4 Search for the Index Set of Unused DoFs

Since the interference channels of the CR receiver \((H^{[0]}, i = 1, ..., K)\) are not aligned, the transformations applied by the precoder vectors of primary transmitters do not change the independent nature of the received signal vectors at this receiver. There-
fore, the CR fusion center can sense the presence or absence of each stream of primary
by an independent binary hypothesis test [15, 10]. To apply hypothesis testing to find
out the indices of the unused DoFs, we assume that the secondary receiver first applies
a sensing vector to the received signal.

2.4.1 Secondary Sensing Vectors

The second phase of the proposed sensing scheme is applied to the received signal at
the fusion center when the presence of spatial hole is detected by the aforementioned
eigenvalue based sensing scheme. In this phase, the fusion center scans the presence
of each DoF of primary network by searching towards the appropriate directions of
the received vectors. This search method can be accomplished by finding a sensing
vector set consisting of \( \sum_{i=1}^{K} d[i] \) vectors \( D[i]^l \) \((i = 1, \cdots, K; l = 1, \cdots, d[i])\). To avoid
the effect of other streams in the sensing of the \( j^{th} \) stream, the direction of the \( j^{th} \)
sensing vector is selected such that it is orthogonal to the summation of all the other
received streams except the \( j^{th} \) stream. Thus, the sensing vector \( D[i]^l \) for finding the
presence of the \( l^{th} \) stream of the \( i^{th} \) primary user is defined as:

\[
D[i]^l \perp \sum_{k=1}^{K} \sum_{j=1}^{d[k]} \frac{\mathbf{H}^{[0k]} \mathbf{V}^{[k]}_j}{\parallel j},
\]

for \( i = 1, \cdots, K \) and \( l = 1, \cdots, d[i] \), where the sensing vector \( D[i]^l \) is of size \( N^{[0]} \times 1 \)
and \( D[i]^l \mathbf{D}^\dagger[i] = 1 \). By this selection, \( D[i]^l \) is orthogonal to the subspace spanned by
all the received vectors except the one corresponding to the \( l^{th} \) stream of the \( i^{th} \)
primary user. In other words, we can find the vector \( D[i]^l \) such that its inner product
with the corresponding received vector is maximized, while \( D[i]^l \notin \text{span}\{H^{[0k]}V^{[k]}_j\} \) for
\( k = 1, \ldots, K, k \neq i \) and \( j = 1, \ldots, d^{[k]}, j \neq l \). By rewriting (2.48) as an inner product, we can find the optimal direction of each sensing vector by formulating the following optimization problem:

\[
\max_{D_i^{[l]}} D_i^{[l]} H^{[0i]} V_i^{[i]} \\
\text{s.t.} \quad D_i^{[l]} H_i^{[0i]} V_i^{[i]} = 0_{1 \times (\sum_{i=1}^{K} d^{[l]} - 1)}, \quad D_i^{[l]} D_i^{[l]} = 1,
\]

where the matrix \( R_i^{[l]} \) of size \( N^{[0]} \times (\sum_{i=1}^{K} d^{[i]} - 1) \) is:

\[
R_i^{[l]} = \begin{bmatrix} H^{[01]} V_1^{[1]}, \ldots, H^{[0v]} V_{l-1}^{[v]}, H^{[0l]} V_{l+1}^{[l]}, \ldots, H^{[0K]} V_K^{[K]} \end{bmatrix}.
\] (2.50)

**Lemma 5.** The sensing vector set is the solution of the optimization problem (2.49) and can be written as:

\[
D_i^{[l]} = \frac{1}{\lambda} \left( H^{[0i]} V_i^{[i]} - C_i^{[l]} \right),
\] (2.51)

where

\[
C_i^{[l]} = R_i^{[l]} \left( R_i^{[l]} R_i^{[l]} \right)^{-1} R_i^{[l]} H^{[0i]} V_i^{[i]},
\] (2.52)

and

\[
\lambda^2 = \left( V_i^{[i]} H^{[0i]} V_i^{[i]} - C_i^{[l]} \right) \left( H^{[0i]} V_i^{[i]} - C_i^{[l]} \right).
\] (2.53)

**Proof.** The first constraint of (2.49) implies that the sensing vector \( D_i^{[l]} \) must be orthogonal onto the subspace of matrix \( R_i^{[l]} \) in \( N^{[0]} \)-dimensional space of the secondary receiver. The orthogonal projection of the desired vector \( H^{[0i]} V_i^{[i]} \) onto subspace \( R_i^{[l]} \) is [88, 5.13]:

\[
C_i^{[l]} = R_i^{[l]} \left( R_i^{[l]} R_i^{[l]} \right)^{-1} R_i^{[l]} H^{[0i]} V_i^{[i]},
\] (2.54)
On the other hand, the objective function of (2.49) is to find an orthogonal vector to $C_l^{[i]}$ with maximum inner product with the desired vector $H^{[0i]}V_l^{[i]}$. Therefore, the normalized vector $D_l^{[i]}$ should be in the direction of the orthogonal projector onto subspace $R_l^{[i]}$ and can be written as:

$$D_l^{[i]} = \frac{1}{\lambda} \left( H^{[0i]}V_l^{[i]} - C_l^{[i]} \right), \quad (2.55)$$

where normalization parameter $\lambda$ is:

$$\lambda^2 = \left( V_l^{[i]*}H^{[0i]*} - C_l^{[i]*} \right) \left( H^{[0i]}V_l^{[i]} - C_l^{[i]} \right). \quad (2.56)$$

**Remark 4** (Feasibility of Sensing). The minimum required number of antennas in the fusion center is equal to the total active DoFs of the primary system. In fact, the rank of matrix $R_l^{[i]}$ must be less than the number of secondary receive antennas, $N^{[0]}$. We should have $N^{[0]} > \sum_{i=1}^{K} d^{[i]} - 1$ antennas to find a vector in null space of $R_l^{[i]}$ to maximize the inner product in the direction of the desired vector, $H^{[0i]}V_l^{[i]}$.

2.4.2 Sensing Criterion and the Probability of False Alarm

Based on the available set of sensing vectors, the fusion center is able to perform the scanning by multiplying the received signal by each sensing vector. If we consider $T$ samples of the received signal, the hypothesis testing problem for the $l^{th}$ stream of the $i^{th}$ primary user can be written as:

$$\mathcal{H}_0 : y_l^{[i]}[n] = \bar{z}[n], \quad n = 0, ..., T - 1 \quad (2.57)$$

$$\mathcal{H}_1 : y_l^{[i]}[n] = \frac{p^{[i]}}{d^{[i]}} D_l^{[i]*} H^{[0i]} \tilde{V}_l^{[i]} X_l^{[i]}[n] + \bar{z}[n], \quad n = 0, ..., T - 1$$
where $\tilde{z}[n]$ is the output noise after the decoder. Here, since we consider normalized sensing vectors in the second constraint of the optimization problem (2.49), the statistics of the noise $\tilde{z}[n]$ after multiplication remain unchanged.

In this scenario, there is an uncertainty about the amplitude of the signal in $\mathcal{H}_1$ hypothesis because of the unknown power allocation scheme in the primary transmitters. In classical detection theory, there are two main approaches to tackle this problem: the Bayesian and the GLRT methods [15, 10]. In the Bayesian method, the unknown parameter can be treated as a random variable with known distribution [15]. The likelihood functions can then be achieved by calculating the marginal probability based on the prior distribution of the unknown parameter. Since the choice of a priori distributions affects the detection performance dramatically and also calculating the marginal distributions is often not tractable, in this research work we adopt the GLRT as a suboptimal detection method [89, Sec. 6.4.2].

In this method, we first estimate the unknown parameters, and then use a likelihood ratio test to make a decision. We assume that the distribution of the signal is complex Gaussian with zero mean and unknown variance. So, if we consider the signal term in $\mathcal{H}_1$, i.e.:

$$s_{l[i]}[n] = \frac{p_{l[i]}}{d_{l[i]}} D_{l[i]}^{[i]} H_{l[0]}^{[i]} V_{l[l]} X_{l[l]}^{[i]}[n], \quad (2.58)$$

and by converting into vector representation, the hypothesis test can be rewritten as:

$$\mathcal{H}_0 : Y_{l[i]} = \tilde{Z}, \quad (2.59)$$
$$\mathcal{H}_1 : Y_{l[i]} = S_{l[i]} + \tilde{Z},$$
where the sample vectors $Y_i^{[i]}$, $S_i^{[i]}$, and $\tilde{Z}$ are of size $T \times 1$.

The general solution for such a hypothesis testing with unknown parameter is to estimate the unknown parameter by the well-known maximum likelihood estimation (MLE) method under $H_1$, i.e.:

$$\hat{\Theta}_1 = \arg\max_{\Theta_1} p(Y_i^{[i]}|H_1, \Theta_1), \quad (2.60)$$

where $p(\cdot)$ denotes the probability density function (PDF). Note again that our assumption is that we have no knowledge of the signal variance, but the noise variance $\sigma_z^2$ is assumed known, i.e. $\tilde{Z} \sim \mathcal{CN}(\mathbf{0}_{T \times 1}, \sigma_z^2 \mathbf{I}_T)$. It is also assumed that $S_i^{[i]}$ and $\tilde{Z}$ are independent and jointly Gaussian.

Based on Neyman-Pearson (NP) theorem [89, Sec. 3.3], for a given PFA, the test statistic that maximizes the probability of detection (PD) is:

$$L_{\text{GLRT}} = \frac{p(Y_i^{[i]}|H_1, \hat{\Theta}_1)}{p(Y_i^{[i]}|H_0)} \gtrless \eta, \quad (2.61)$$

where there is no unknown parameter in hypothesis $H_0$, and we have:

$$p(Y_i^{[i]}|H_0) = \prod_{n=0}^{T-1} \frac{1}{2\pi\sigma_z^2} \exp \left[ -\frac{1}{2\sigma_z^2} ||y_i^{[i]}[n]||^2 \right] \quad (2.62)$$

Hence:

$$\ln p(Y_i^{[i]}|H_0) = -T \ln (2\pi\sigma_z^2) - \frac{1}{2\sigma_z^2} Y_i^{[i]\dagger} Y_i^{[i]}. \quad (2.63)$$

On the other hand, the likelihood function under $H_1$ with the unknown parameter $\sigma_s$ can be written as:

$$p(Y_i^{[i]}|H_1, \sigma_s^2) = \prod_{n=0}^{T-1} \frac{1}{2\pi(\sigma_s^2 + \sigma_z^2)} \exp \left[ -\frac{1}{2(\sigma_s^2 + \sigma_z^2)} ||y_i^{[i]}[n]||^2 \right]. \quad (2.64)$$
Taking logarithm from both sides, we get:

\[ \ln p(Y_i[^{[i]}]|H_1, \sigma_s^2) = -T \ln \left( 2(\sigma_s^2 + \sigma_z^2) \right) - \frac{1}{2(\sigma_s^2 + \sigma_z^2)} Y_i[^{[i]}]^\dagger Y_i[^{[i]}]. \]  

(2.65)

The first derivative of the log-likelihood function with respect to \( \sigma_s^2 \) is:

\[ \frac{\partial \ln p(Y_i[^{[i]}]|H_1, \sigma_s^2)}{\partial \sigma_s^2} = -\frac{T}{\sigma_s^2 + \sigma_z^2} + \frac{1}{2(\sigma_s^2 + \sigma_z^2)^2} Y_i[^{[i]}]^\dagger Y_i[^{[i]}]. \]  

(2.66)

Therefore, the MLE of the unknown parameter \( \sigma_s^2 \) can be obtained by finding the root of the first derivative as:

\[ \hat{\sigma}_s^2 = \frac{Y_i[^{[i]}]^\dagger Y_i[^{[i]}]}{2T} - \sigma_z^2. \]  

(2.67)

By substituting (2.67) in (2.65), we have:

\[ \ln p(Y_i[^{[i]}]|H_1, \hat{\sigma}_s^2) = -T \ln \left( \frac{\pi Y_i[^{[i]}]^\dagger Y_i[^{[i]}]}{T} \right) - T. \]  

(2.68)

By plugging (2.63) and (2.68) into (2.61), the log-GLRT statistics becomes:

\[ \ln L_{\text{GLRT}}(Y_i[^{[i]}]) = T \ln \left( \frac{2T\sigma_z^2}{Y_i[^{[i]}]^\dagger Y_i[^{[i]}]} \right) + \frac{1}{2\sigma_z^2} Y_i[^{[i]}]^\dagger Y_i[^{[i]}] - T, \]  

(2.69)

Therefore, the final test can be written as:

\[ L_{\text{GLRT}}(Y_i[^{[i]}]) = \left( \frac{2T\sigma_z^2}{Y_i[^{[i]}]^\dagger Y_i[^{[i]}]} \right)^T \exp \left( \frac{1}{2\sigma_z^2} Y_i[^{[i]}]^\dagger Y_i[^{[i]}] - T \right)^\frac{\mathcal{H}_1}{\mathcal{H}_0} \eta. \]  

(2.70)

We can define the new threshold for the test statistics as:

\[ T(Y_i[^{[i]}]) = \frac{1}{Y_i[^{[i]}]^\dagger Y_i[^{[i]}]} \exp \left( \frac{Y_i[^{[i]}]^\dagger Y_i[^{[i]}]}{2T\sigma_z^2} \right)^\frac{\mathcal{H}_1}{\mathcal{H}_0} \frac{\eta^\frac{1}{2}}{2T\sigma_z^2} e^{\frac{1}{2}} = \eta'. \]  

(2.71)

Similar to the NP theorem [89, Sec. 3.3], the threshold \( \eta' \) can be obtained for an arbitrary PFA, i.e.:

\[ P_{\text{FA}} = p(T(Y_i[^{[i]}]) > \eta'; \mathcal{H}_0). \]  

(2.72)
For finding this probability, the distribution of the random variable \( T(Y^i_l) \in \mathbb{R} \) in (2.71) should be derived. More precisely, the PFA for a predefined threshold can be written as:

\[
P_{FA} = 1 - F_{\vartheta^i_l}(\eta')
\]  

(2.73)

where \( F_{\vartheta^i_l} \) is the CDF of \( \vartheta^i_l = \frac{1}{\sigma^2} Y^i_l Y^{i*}_l, \vartheta^i_l \sim \chi^2(T) \), and \( \chi^2(T) \) is the central Chi-squared distribution with \( T \) degrees of freedom. We can now rewrite \( T(Y^i_l) \) in (2.71) as a function of the new variable \( \vartheta^i_l \) as:

\[
T(Y^i_l) = g(\vartheta^i_l) = \frac{1}{\sigma^2} \exp \left( \frac{\vartheta^i_l}{2T} \right).
\]  

(2.74)

The function \( g \) is a two-to-one function with a minimum at \( \vartheta^i_l = 2T\sigma^2_z \). An approximate inverse function of \( g \) can be derived as [90, 91]:

\[
g^{-1}(\eta') = \begin{cases} 
-(2T)W_0 \left( -\frac{1}{(2T\sigma^2_z)\eta'} \right) & \text{for } 0 < \vartheta^i_l < (2T) \\
-(2T)W_{-1} \left( \frac{-1}{(2T\sigma^2_z)\eta'} \right) & \text{for } \vartheta^i_l \geq (2T)
\end{cases}
\]  

(2.75)

where \( W_0(\cdot) \) and \( W_{-1}(\cdot) \) are the lower and upper branches of Lambert W function, respectively, where the general form of \( W(z) \) is defined by:

\[
F_z = W(z)e^{W(z)}.
\]  

(2.76)

The PFA can be finally written as:

\[
P_{FA} = 1 - F_{\vartheta^i_l}(\eta') = 1 - p \left( g^{-1}_{\text{left}}(\eta') \leq \vartheta^i_l < g^{-1}_{\text{right}}(\eta') \right) 
\]

(2.77)

\[
= 1 - \mathcal{P} \left[ -(2T)W_{-1} \left( \frac{-1}{(2T\sigma^2_z)\eta'} \right), T \right] + \mathcal{P} \left[ -(2T)W_0 \left( \frac{-1}{(2T\sigma^2_z)\eta'} \right), T \right],
\]

where \( \mathcal{P} [\cdot, \cdot] \) stands for the regularized Gamma function which is defined as:

\[
\mathcal{P}[x, k] = \frac{\gamma(x, k)}{\Gamma(x)}.
\]  

(2.78)
where $\Gamma(x)$ and $\gamma(x, k)$ are the Gamma and the lower incomplete Gamma functions, respectively. Note that each test statistic is evaluated to detect a certain data stream. Therefore, when there are more than one unused DoFs, the PFA analysis will be the same as each direction is considered independently.

2.4.3 Advantages of Proposed Sensing Method

The overall sensing scheme is summarized in Alg. 1. The DoF index search (second step) is performed only when the eigenvalue based method (first step) detects the presence of unused DoFs. There are two advantages in dividing the sensing process into two steps. First, the overall complexity of the sensing process is significantly reduced especially because there is no need to perform the complex DoF index search if no unused DoF is detected in the first step. The second reason for performing the eigenvalue based sensing in the first step is the ability of this method to sense the presence of spatial holes without the need for primary CSI. Therefore, the primary CSI is only acquired when the first sensing step flags the presence of spatial hole(s).

The complexity order of the eigenvalue based sensing is $\mathcal{O}(N^{[0]^2})$ multiplications for calculating the covariance matrix $R_Y$, while the eigenvalues can be calculated by at most $\mathcal{O}(N^{[0]^3})$ calculation in each sensing frame. On the other hand, the second phase of the proposed sensing algorithm requires several matrix operations. The number of multiplications for finding the sample vectors of all data streams is in the order of $\mathcal{O}(D^4)$ where $D = \sum_{i=1}^{K} d[i]^K$. If we assume the number of antennas at the secondary receiver (fusion center) is equal to the total number of data streams,
1: Calculate the sensing covariance matrix $R_Y$ using (2.34)

2: Calculate the minimum eigenvalue of $R_Y$

3: while $\lambda_{\text{min}}(R_Y) < \eta$

4:   for $i = 1 : K$

5:       for $l = 1 : d[i]$

6:           Calculate the sensing vector $D_i^{[i]}$ using (2.51)

7:       Calculate the sample vector $Y_i^{[i]}$

8:       Calculate the test statistics $T(Y_i^{[i]})$ using (2.71)

9:       if $T(Y_i^{[i]}) < \eta'$

10:          The $l^{th}$ data stream of the $i^{th}$ user is unused

11:     end if

12:   end for

13: end for

14: Calculate the sensing covariance matrix $R_Y$ using (2.34)

15: Calculate the minimum eigenvalue of $R_Y$

16: end while

**Algorithm 1: Proposed sensing algorithm**

the order of required multiplications for the DoF index search is $O(N^{[n]})$. The test statistics $T(Y_i^{[i]})$ is also required to be calculated which is an exponential function. Moreover, for finding the threshold value in the proposed eigenvalue based method, the primary CSI is not required (only the covariance matrix $\Sigma$ should be calculated).
2.5 Simulation Results of Cognitive Interference Alignment Networks

We consider a 3-user IA system with two antennas in each node as the primary network and a secondary MIMO link as shown previously in Fig. 2.2. It is assumed that the design of IA precoders at the primary is done with the distributed numerical approach presented in [75] with 20 iterations. The secondary system applies the transmission scheme proposed in this work. We assume that all the channel links are zero-mean unit-variance complex Gaussian distributed. We further assume that all the receive noise are zero mean complex Gaussian with unit variance. It is also considered that the same transmit power of 10 dBW is used for both the primary and secondary systems in all scenarios.

2.5.1 Number of Secondary Antennas and Feasibility of CR

Fig. 2.3 shows the effect of the number of secondary antennas on the total interference leakage (IL) as a function of the SNR at the active primary receivers when there is only one inactive primary data stream. This figure also illustrates the interference from the primary transmitters to the secondary receiver for different number of secondary antennas. As seen, when the number of transmit antennas at the secondary is equal to or more than the total number of primary streams (in the figure, we have considered $N^{[0]} = 3$), the total interference imposed to the primary network due to secondary transmission is less than the total interference at each primary receiver caused by other primary transmitters when all the primary pairs are active and there is no unused DoF. This means that applying our secondary transmission scheme does not
increase the total interference power at the primary receivers. On the other hand, in the case of two transmit antennas at the secondary \(N[0] = 2\), the interference imposed to the primary is significantly larger, and the CR transmission is not feasible. This shows the significance of the inequality proved in Lemma 1 for the minimum required number of antennas at the secondary transmitter.

The case of two inactive streams is illustrated in Fig. 2.4. We assume that the secondary network uses both of the spatial holes for its own transmission. This figure shows that the proposed method can be equally applied for opportunistic use of more than one spatial hole. In this case, since we consider two streams for the secondary network, the interference to the CR receiver increases when the number of antennas at CR transmitter is less than three. Since the interference leakage in the active primary receiver is almost zero for the three transmit antennas, we have not shown this curve in the figure. On the other hand, the fundamental requirement for the secondary network is to be able to communicate at reasonable rate in the primary spectrum band. To see this, we have simulated the throughput of the secondary system in Fig. 2.5. This figure also shows the sum-rate performance of the primary system for different secondary antenna configurations. Based on this figure, when \(M[0] = 3\), i.e. when the condition on the minimum number of antenna is satisfied, the secondary transmission, when primary transmitter one is off, does not affect the average sum-rate of pairs 2 and 3 of the primary network (i.e. \(R_2 + R_3\)). However, this is not the case when \(M[0] = 2\). In this case, i.e. when the number of transmit antennas at the secondary is less than the total DoFs of the primary, the sum-rate
Figure 2.3: Interference power at the primary and secondary receivers for different scenarios when there is one inactive primary user.

$(R_2 + R_3)$ of the primary degrades significantly due to the non-negligible secondary interference. This figure also shows that by increasing the number of receive antennas at the CR, its average rate increases significantly due to the secondary decoder.

In Fig. 2.6, we have simulated the sum rates for the case of two inactive primary streams, while the CR utilizes both available spatial holes for its own transmission. This figure shows that the secondary network does not affect the achievable rate of the primary system whenever the condition of Lemma 1 is met. On the other hand, the CR network is able to transmit two data streams and achieve a usable rate.

Fig. 2.7 illustrates the average rate of the secondary as well as the average sum-rate of the primary as a function of the number of secondary receive antennas.
As shown while the number of secondary receive antennas does not have a significant effect on the sum-rate of the primary, the average secondary rate is always an increasing function of its number of receive antennas. This is in agreement with the result of Lemma 3. This figure also shows that the statement of the Lemma 3 is still valid for the case of more than one secondary data stream, and the self-interference does not change the behavior of average SINR with respect to number of antennas at the secondary. The figure also illustrates that after some points, increasing the number of secondary receive antennas has a small effect on its average rate. For example, according to Fig. 2.7, increasing the number of antennas from 8 to 10 increases the average rate of the secondary by only 0.4 bps/Hz.
2.5.2 Performance of Fast DoF Sensing Method

Fig. 2.8 shows the capability of the proposed eigenvalue based sensing scheme in detecting the presence of unused DoFs. We set the number of samples for producing the correlation matrix to $L = 30$ to analyze the effect of number of samples on the performance of sensing method. The figure shows the PFA, and its theoretical upper and lower bounds as well as the correct detection probability as a function of threshold $\eta$ for two different values of smoothing factor $T$. As shown, at sufficiently large number of samples, the PFA suddenly drops while the PD is still high. Note that the theoretical upper bound of the PFA is fairly close to the Monte Carlo sim-
Figure 2.6: Average sum-rate of the primary system versus SNR before and after CR transmission when there are two inactive primary users. Figure also shows the average rate of the secondary for different antenna configurations.

ulation results. This figure also shows that the PFA decreases when there are more than one spatial holes. To further demonstrate the capability of the proposed DoF sensing method in detecting the presence (or absence) of spatial hole(s), we evaluate its receiver operating characteristic (ROC) for different smoothing factors, $T$, and the number of secondary receive antennas, $N^{[0]}$. Fig. 2.9 shows the ROC of the proposed fast sensing method, and as seen, the ROC becomes sharper by increasing $T$ and $N^{[0]}$. The plot also shows that, especially for $N^{[0]} = 4$, the probabilities of detection and false alarm are very close to one and zero, respectively. This means that the fast sensing method can reliably detect the presence and absence of spatial hole(s). Therefore, using it as a precursor step before performing DoF index search
can significantly reduce the computational complexity at no or small hit in sensing performance.

2.5.3 Performance of DoF Index Search Method

The ROC performance results of the proposed searching method is shown in Fig. 2.10. In this scheme, the number of sensing antennas is an important factor in finding the unused DoFs. It can be seen that, by increasing the number of secondary receive antennas, the PD for a specific PFA significantly increases, especially at a low PFA. As also shown in the figure, the experimental results matches perfectly with our theoretical analysis. The number of samples in the proposed GLRT detection is an important factor in the performance. As seen, the performance of the sensing scheme
Figure 2.8: PFA, its lower and upper bounds, and the PD versus threshold for null space sensing.

significantly increases with number of samples in each test. It can bee seen that for a PFA less than 0.3, we can find a threshold that achieves a PD higher than 0.5 to make the scheme practically feasible.
Figure 2.9: ROC curves of the null space sensing method for different smoothing factors and number of secondary receive antennas.

Figure 2.10: ROC curves of the DoF index search method for different smoothing factors and number of secondary receive antennas.
CHAPTER III
COGNITIVE SPATIAL MODULATION

In this chapter, we propose a novel overlay scheme in which the CR transmitter utilizes spatial modulation while assisting the primary network in transmitting its information using amplify-and-forward (AF) relaying. In the proposed scheme, the secondary network transmits its own information in the space domain without causing any interference to the primary network. At the same time, the APM symbols of the primary transmitter is amplified and retransmitted via the active antenna of the secondary transmitter, which will then be detected at the primary receiver. We firstly introduce the proposed scenario and the mathematical formulation of the scheme. Then the optimal detection scheme and its corresponding pair-wise error is provided. Then the average symbol error probability is closely calculated in the next section and the results for high SNR regime is also approximated. The system design parameters of the proposed model and the simulation results are provided in the last two section of this chapter.

3.1 System Model

The overlay CR network assists the primary source to transmit its information to the primary receiver [92]. The secondary transmitter, that serves as an AF relay for the primary transmission, operates in two phases. In the first transmission interval,
the primary transmitter broadcasts APM symbols, while the secondary transmitter is silent, and in the second transmission interval, the secondary transmitter transmits its own information, while relaying the APM signal of the primary. Even though the proposed scheme works regardless of the presence or absence of a direct link between the primary transmitter and receiver, for the sake of tractability of the analytical derivations, it is assumed that the direct link is broken. This assumption is also suitable for a primary receiver that is far from the primary transmitter, e.g. a cell edge user, which experiences very low SNRs [93]. Therefore, the secondary network can effectively connect the primary transmitter and receiver while transmitting its own information over the primary’s licensed spectrum band. Although, in this case, the primary transmission depends on the presence of the secondary network, the primary transmitter and receiver cannot otherwise communicate without the assistance of the secondary. On the other hand, the primary performance results obtained with this assumption can be considered as a benchmark for the worst case scenario, and the presence of the direct primary link always results in a better performance.

It is further assumed that the primary network consists of a single-antenna transmitter (S) and multiple-antenna receiver (D). As shown in Fig. 1, the secondary transmitter (R) transmits with multi-antenna using SM but receives with only a single antenna. This assumption is important as it enables the secondary transmitter to retain a simple hardware architecture that has one transmit/receive radio frequency

1Alternatively, the primary can include a multi-antenna transmitter, but a single APM symbol is transmitted in each transmission interval.
We also assume that the channel state information (CSI) is available at both the primary and the secondary receivers. Although acquiring full CSI at the primary and secondary systems is difficult in practice, there are a number of ways to make this possible. One way to acquire such information is transmitting pilot data from primary transmitter. Another practical way to obtain CSI for efficient CR operation using cognitive pilot channel (CPC) has been introduced in [73], and currently standardization bodies are considering using such channels to broadcast useful information of primary to available secondary networks for opportunistic or underlay utilization of the spectrum [74]. Also, considering the feedback channels as an alternative way to acquire CSI in CR networks has been studied in [52] and [49].

3.1.1 Primary Transmission

As shown in Fig. 3.1, the secondary transmitter (R) receives the primary APM symbol in the first transmission interval from the primary transmitter (S), and this signal can be written as:

\[ y_R = \sqrt{P_S} h_{SR} x + n_R, \]  

(3.1)

where \( P_S \) is the transmission power of primary transmitter, \( h_{SR} \) is the channel gain between the primary transmitter and the secondary transmitter, which is assumed to be a zero-mean complex Gaussian random variable with variance \( \sigma^2_{SR} \), \( x \) is the primary APM symbol selected from a constellation space \( \chi^M \), where \( M \) is the constellation size, and \( n_R \) is the zero-mean complex Gaussian noise at the secondary transmitter.
with the variance $\sigma_R^2$. In writing (3.1), it is assumed that the secondary transmitter receives the primary signal using one of its antennas, or otherwise, does not combine the signal observations received over all its antennas.

3.1.2 CR Transmission

In the second transmission interval, the initial role of the secondary transmitter is to amplify the received signal and retransmit it to the primary receiver (D) using its $N_R$ antennas. On the other hand, the secondary transmitter has its own information to be sent to the corresponding secondary receiver (C). Therefore, in our scheme, the secondary transmitter utilizes an SM transmission in such a way that the primary symbols are relayed by the secondary transmitter, while the secondary symbols are also transmitted simultaneously using the antenna indices. We denote by $s$ the index
of the transmitting antenna at the secondary transmitter. Therefore, the received
signal at the primary receiver can be written as:

\[ y_D = \sqrt{P_R} \alpha h_{RD}^{(s)} y_R + n_D, \quad (3.2) \]

and at the secondary receiver as:

\[ y_C = \sqrt{P_R} \alpha h_{RC}^{(s)} y_R + n_C, \quad (3.3) \]

where \( P_R \) is the transmission power of the secondary transmitter, \( \alpha \) is the power
scaling factor, channel vectors \( h_{RD}^{(s)} \) and \( h_{RC}^{(s)} \) are the \( s \)th columns of the \( N_D \times N_R \)
and \( N_C \times N_R \) channel matrices \( H_{RD} \) and \( H_{RC} \), respectively. All the elements of the
channel vectors are assumed to be independent zero-mean complex Gaussian random
variables with variances \( \sigma_{RD}^2 \) and \( \sigma_{RC}^2 \), respectively. Also, \( n_D \) and \( n_C \) are the zero-
mean complex Gaussian noises with variances \( \sigma_{D}^2 I_{N_D} \) and \( \sigma_{C}^2 I_{N_C} \), where \( I \) is identity
matrix.

Assuming fixed gain AF relaying, the power scaling factor at the secondary
transmitter can be written as:

\[ \alpha = \frac{1}{\sqrt{\mathbb{E}[|y_R|^2]}} = \frac{1}{\sqrt{P_S \sigma_{SR}^2 \sigma_\chi^2 + \sigma_R^2}}, \quad (3.4) \]

where \( \mathbb{E}[x] \) denotes the expected value of \( x \), and \( \sigma_\chi^2 \) is the average power of the
APM constellation of size \( M \). Without loss of generality, we assume that \( \sigma_\chi^2 = 1 \).

By substituting (3.1) into (3.2) and (3.3), the received signals at the primary and
secondary receivers can be rewritten as:

\[ y_D = \sqrt{P_S P_R} \alpha h_{SR}^{(s)} h_{RD}^{(s)} x + \sqrt{P_R} \alpha h_{RD}^{(s)} n_R + n_D, \quad (3.5) \]
and

\[ y_C = \sqrt{P_S P_R} \alpha h_{SR}^{(s)} x + \sqrt{P_R} \alpha h_{RD}^{(s)} n_R + n_C. \]  

(3.6)

It is worth mentioning that the DF scheme is potentially applicable in the proposed scenario instead of AF as provided in [55, 56]. For the case of DF scheme, the amplified noise term does not appear at the primary receiver while the channel \( h_{SR} \) should be known at the secondary transmitter. In the next section, we introduce optimal detection schemes in both receivers.

### 3.2 Optimal Detection

Although the primary receiver does not need to detect the secondary transmitting antenna index, the channel information for the correct detection of the primary APM symbols is required at its receiver. So, the primary receiver must know the index of the secondary transmitting antenna. We consider a joint ML detection at the primary receiver to identify the transmitted APM symbols as well as the index of the secondary transmitting antenna. We also consider a joint ML detection at the secondary receiver. The joint detection reveals the information of both domains at the receivers. This can be considered as a security thread to the primary and the secondary systems. Privacy of the primary and secondary users in CR networks has been recently studied in [94] and [95]. A practical solution to this is to apply encryption and cyphering schemes at higher layers to ensure the privacy of the primary and secondary systems.
3.2.1 Primary and Secondary Optimal Detector

The ML detection rule for the primary receiver can be written as:

\[
\tilde{p} = \arg \max_{s,p} f \left( y_D | h_{RD}^{(s)}, h_{SR}, x^{(p)} \right),
\]  

(3.7)

where \( f \left( y_D | h_{RD}^{(s)}, h_{SR}, x^{(p)} \right) \) is the conditional probability density function (PDF) of the signal at the primary receiver when the \( p \)th APM symbol via the \( s \)th secondary antenna is transmitted. Considering the Gaussian distribution for the received signal and with some manipulations, one can derive the optimal detection rule at the primary receiver as [96]:

\[
\tilde{p} = \arg \max_{s,p} \Re \left\{ \left( y_D - \frac{\sqrt{P_S} P_R}{2} \alpha h_{SR} h_{RD}^{(s)} x^{(p)} \right)^H h_{RD}^{(s)} h_{SR} x^{(p)} \right\},
\]  

(3.8)

where \( \Re (\cdot) \) denotes the real part. The detection process at the secondary receiver is quite similar to the joint optimal detector at the primary receiver such that only the antenna index \( s \) is determined. Therefore, the detection rule at the secondary receiver can be written as [96]:

\[
\tilde{s} = \arg \max_{s,p} \Re \left\{ \left( y_C - \frac{\sqrt{P_S} P_R}{2} \alpha h_{SR} h_{RC}^{(s)} x^{(p)} \right)^H h_{RC}^{(s)} h_{SR} x^{(p)} \right\}.
\]  

(3.9)

As can be seen in (8) and (9), the detectors do not require the channel knowledge of the first and second link of relaying separately, and estimating the multiplication of the two channel gains is sufficient for detection.

3.2.2 Pair-Wise Error Probability at the Primary Receiver

We separate symbol PEP into two types. The first type of error happens when the joint ML detector cannot detect both the APM symbol and the antenna index
correctly. The second type includes cases when the antenna index is detected correctly while the APM symbol is detected incorrectly. Assuming that the $p^{th}$ symbol of the primary is transmitted via the $s^{th}$ antenna at the secondary transmitter, the union bound on the probability of symbol error at the primary receiver can be written as [96]:

$$P_{\text{Primary}} \leq \frac{1}{N_R M} \sum_{s=1}^{N_R} \sum_{p=1}^{M} \text{PEP}_1(p, p', s, s') + \frac{1}{N_R M} \sum_{s=1}^{N_R} \sum_{p=1}^{M} \text{PEP}_2(p, p', s),$$

(3.10)

where the $\text{PEP}_1$ term, corresponding to the joint errors, can be easily derived from (8) as:

$$\text{PEP}_1(p, p', s, s') = Q\left(\frac{P_S P_R \alpha^2 |h_{SR}|^2 \left|h_{RD}^{(s)} x(p) - h_{RD}^{(s')} x(p')\right|^2}{2 P_R \alpha^2 \left|h_{RD}^{(s)}\right|^2 \sigma_R^2 + \sigma_D^2}\right),$$

(3.11)

and the $\text{PEP}_2$ term, corresponding to the APM domain, can be written as:

$$\text{PEP}_2(p, p', s) = Q\left(\frac{P_S P_R \alpha^2 |h_{SR}|^2 \left|h_{RD}^{(s)}\right|^2 \left|x(p) - x(p')\right|^2}{2 P_R \alpha^2 \left|h_{RD}^{(s)}\right|^2 \sigma_R^2 + \sigma_D^2}\right),$$

(3.12)

where $p'$ and $s'$ are the detected indices of the primary APM symbol and the secondary antenna at the primary receiver, respectively.
3.2.3 Pair-Wise Error Probability at the Secondary Receiver

Similarly, the union bound of the error at the secondary receiver can be written as:

$$
\mathcal{P}_{\text{Secondary}} \leq \frac{1}{N_R M} \sum_{s=1}^{N_R} \sum_{s' \neq s, p=1}^{M} \sum_{p'=1}^{M} \text{PEP}_3(s, s', p, p') 
$$

$$
+ \frac{1}{N_R M} \sum_{s=1}^{N_R} \sum_{s' \neq s, p'=1}^{M} \sum_{p=1}^{M} \text{PEP}_4(s, s', p),
$$

where the joint error term is calculated from (9) as:

$$
\text{PEP}_3(s, s', p, p') = Q \left( \sqrt{\frac{P_S P_R \alpha^2 |h_{SR}|^2 |h_{RC}^{(s)} x^{(p)} - h_{RC}^{(s')} x^{(p')}|^2}{P_R \alpha^2 |h_{RC}^{(s)}|^2 \sigma_R^2 + \sigma_C^2}} \right),
$$

and the term corresponding to the spatial domain is:

$$
\text{PEP}_4(s, s', p) = Q \left( \sqrt{\frac{P_S P_R \alpha^2 |h_{SR}|^2 |h_{RC}^{(s)} - h_{RC}^{(s')}|^2 |x^{(p)}|^2}{P_R \alpha^2 |h_{RC}^{(s)}|^2 \sigma_R^2 + \sigma_C^2}} \right).
$$

In the next section, we derive the average error probability at the primary and secondary receivers.

3.3 Average Symbol Error Rate

In this section, we derive a closed-form equation for the average symbol error probability based on the instantaneous probability of error. Since the analysis of the error for multiple receive antennas is cumbersome, hereafter, we consider the case of single-antenna receivers. Later in Section VII, we also numerically simulate the performance of the proposed scheme in the case of multi-antenna receivers.
It is possible to obtain the average error performance by averaging over the instantaneous pair-wise error in the union bound [6]. In this work, however, we derive the ASER using cumulative density function (CDF) approach [97]. Using this approach, one can write the average PEP as:

$$PEP_i = \int_0^\infty -PEP'_i(a_i, \gamma_i) F_{\gamma_i}(\gamma_i) \, d\gamma_i,$$  \hspace{0.5cm} (3.16)

for \( i = 1, 2, 3, 4 \), where \( PEP'_i(a_i, \gamma_i) = (d/d\gamma_i) PEP_i(a_i, \gamma_i) \) denotes the derivative of the PEP function defined as:

$$PEP_i(a_i, \gamma_i) = Q\left(\sqrt{a_i \gamma_i}\right).$$  \hspace{0.5cm} (3.17)

For deriving an upperbound on the ASER, we need to calculate the average PEPs in Section III. By substituting the first derivative of (3.17) into (3.16), we have:

$$PEP_i = \frac{1}{2} \sqrt{\frac{a_i}{2\pi}} \int_0^\infty e^{-\frac{a_i \gamma_i}{2}} \sqrt{\gamma_i} F_{\gamma_i}(\gamma_i) \, d\gamma_i.$$  \hspace{0.5cm} (3.18)

Our aim is to derive the CDF of \( \gamma_i \) to calculate the ASER based on (3.18). For this purpose, we define the random variables \( \gamma_i \ (i = 1, \ldots, 4) \) as follows:

$$\gamma_1 = \frac{|h_{SR}|^2 \left|h_{RD}^{(s)} x^{(p)} - h_{RD}^{(s')} x^{(p')}\right|^2}{\left|h_{RD}^{(s)}\right|^2 + c},$$  \hspace{0.5cm} (3.19)

$$\gamma_2 = \frac{|h_{SR}|^2 \left|h_{RD}^{(s)}\right|^2}{\left|h_{RD}^{(s)}\right|^2 + c},$$  \hspace{0.5cm} (3.20)

$$\gamma_3 = \frac{|h_{SR}|^2 \left|h_{RC}^{(s)} x^{(p)} - h_{RC}^{(s')} x^{(p')}\right|^2}{\left|h_{RC}^{(s)}\right|^2 + c'},$$  \hspace{0.5cm} (3.21)

$$\gamma_4 = \frac{|h_{SR}|^2 \left|h_{RC}^{(s)}\right|^2}{\left|h_{RC}^{(s)}\right|^2 + c'}.$$  \hspace{0.5cm} (3.22)
where \( c = \frac{\sigma_D^2}{P_{\alpha^2}\sigma_R^2} \) and \( c' = \frac{\sigma_C^2}{P_{\alpha^2}\sigma_R^2} \). The APM constellation design for the SM with ML detector has been studied in [34]. It is shown that although conventional quadrature amplitude modulation (QAM) outperforms PSK, SM-PSK is generally performs better than SM-QAM [96]. So, besides the fact that the analysis of SER for the case of PSK modulation is simpler and more tractable compared to that of the QAM, it appears to be a more suitable candidate for the proposed CR scenario. Therefore, we only investigate the PSK modulation in this work. The analysis of other types of APM schemes is left for a future work. By this assumption and considering (3.17), we have \( a_1 = a_3 = a_4 = \frac{P_S}{2\sigma_R^2} \) and \( a_2(p,p') = \frac{P_S|z(p) - z(p')|^2}{2\sigma_R^2} \). Without loss of generality, we also assume equal noise powers at both receivers, i.e. \( \sigma_D^2 = \sigma_C^2 \), and then \( c = c' \).

Considering that \( \gamma_i \)'s \( (i = 1, \ldots, 4) \) include the ratio of two correlated random variables, calculating their CDFs is not straightforward. Therefore, we first approximate the four CDFs, and then, calculate the average PEPs. By achieving the average PEPs, the ASER can be derived subsequently.

Let \( Z_1, Z_2 \geq 0 \) be two, not necessarily identically distributed, Exponential random variables. The joint PDF of \( Z_1 \) and \( Z_2 \) can be written as [98, eq 1]:

\[
f_{Z_1,Z_2}(z_1,z_2) = \frac{1}{\mu_1\mu_2(1-\rho)} \times \exp \left[-\frac{1}{1-\rho}\left(\frac{z_1}{\mu_1} + \frac{z_2}{\mu_2}\right)\right] I_0 \left[\frac{2\sqrt{\rho z_1 z_2}}{(1-\rho)\sqrt{\mu_1 \mu_2}}\right],
\]

where \( \mu_l = \mathbf{E}(Z_l) > 0 \) for \( l = 1, 2 \) and \( I_n(\cdot) \) is the \( n^{th} \)-order modified Bessel function of the first kind, and \( \rho \) is the correlation coefficient between \( Z_1 \) and \( Z_2 \).
Theorem 1. For two Exponentially distributed random variables $Z_1$ and $Z_2$, and a real constant $C$, the PDF of $Z = \frac{Z_1}{Z_2 + C}$ is:

$$f_Z(z) = \frac{1}{\mu_1\mu_2 (1 - \rho)} \exp \left[ \frac{C}{2} \left( s(z) - \beta(z) - \frac{2}{(1 - \rho) \mu_1} \right) \right] \times \left( \frac{s(z)}{\beta(z)^3} + C \left( -\frac{1}{2\beta(z)} + \frac{s(z)}{2\beta(z)^2} + \frac{1}{\beta(z) \mu_1 \mu_2 (1 - \rho)} \right) \right],$$

(3.24)

where $\mu_l = \mathbf{E}(Z_l) > 0$ ($l = 1, 2$), $\rho$ is the correlation coefficient between $Z_1$ and $Z_2$, $s(z) = \frac{(\mu_1 + \mu_2 z)}{\mu_1 \mu_2 (1 - \rho)}$, and

$$\beta(z) = \frac{1}{\mu_1 \mu_2 (1 - \rho)} \sqrt{\mu_2^2 z^2 + 2 \mu_1 \mu_2 (1 - 2 \rho) z + \mu_1^2}.$$  

(3.25)

Proof. See Appendix C.

To find the CDF of the ratio $Z$, we need to integrate over the PDF $f_Z(z)$ in (3.24), which is generally cumbersome. Instead, we derive the CDF in a special case which suffices for our problem. To this end, we first derive the correlation coefficients of $\gamma_1$, $\gamma_3$ and $\gamma_4$ for the case of PSK transmission at the primary.

Lemma 6. Given two independent complex Gaussian random variables $Y$ and $W$ with variances $\sigma_Y^2$ and $\sigma_W^2$, respectively, complex constants $x^{(p)}$, $x^{(p')}$, and real constant $C$, the correlation coefficient between $X_1 = |Y x^{(p)} - W x^{(p')}|^2$ and $X_2 = |Y|^2 + C$ is:

$$\rho_{X_1, X_2} = \frac{|x^{(p)}|^2 \sigma_Y^4}{|x^{(p)}|^2 \sigma_Y^4 + |x^{(p')}|^2 \sigma_W^2 \sigma_Y^2}.$$  

(3.26)

Proof. See Appendix D. 

68
Note that for the special case of equal-variance $Y$ and $W$, i.e. $\sigma_Y^2 = \sigma_W^2$, the correlation coefficient can be simplified as:

$$\rho = \frac{\left| x^{(p)} \right|^2}{\left| x^{(p)} \right|^2 + \left| x^{(p')} \right|^2}, \quad (3.27)$$

In our problem, $Y$ and $W$ correspond to the elements of the channel vectors $h_{RD}$ and $h_{RC}$. As we consider i.i.d. channels and due to the assumption of PSK transmission at the primary, using Lemma 6, it turns out that the correlation coefficient defined in (3.26) is $\rho = \frac{1}{2}$. In the following theorem, we use the result of Lemma 6 to find a closed-form CDF of the ratio of two correlated Exponential random variables in the special case of $C = 0$.

**Theorem 2.** For two, not necessarily identically distributed, Exponential random variables $Z_1$ and $Z_2$ with correlation coefficient $\frac{1}{2}$, the CDF of $Z' = \frac{Z_1}{Z_2}$ is:

$$F_{Z'}(z) = \frac{1}{2} \left( 1 - \frac{\mu_1 - \mu_2 z}{\sqrt{\mu_2^2 z^2 + \mu_1^2}} \right), \quad (3.28)$$

where $\mu_l = E(Z_l) > 0$ ($l = 1, 2$).

**Proof.** By substituting $C = 0$ and $\rho = \frac{1}{2}$ in (3.24) of Theorem 1, the PDF of $Z'$ is simplified to:

$$f_{Z'}(z) = \frac{\mu_1 \mu_2 (\mu_1 + \mu_2 z)}{2(\mu_2^2 z^2 + \mu_1^2)^2}, \quad (3.29)$$

For calculating the CDF of $Z'$, we can simply integrate over the PDF, i.e., $F_{Z'}(z) = \int_0^z f_{Z'}(y) dy$. This integral is easily derived using [99, eq 2.264] and yields (3.28).

For finding the CDF of $\gamma_i$ for $i = 1, 3$ and 4, we also need to derive the CDF of product of the ratio of two correlated random variables from Theorem 2 and another
Exponentially distributed random variable, which is accomplished in the following lemma.

**Lemma 7.** Let $Z_1$ and $Z_2$ be two correlated Exponentially distributed random variables with correlation coefficient $\frac{1}{2}$ and $U$ be an independent Exponential random variable. The CDF of $V = U \frac{Z_1}{Z_2}$ is:

$$F_V(v) = \frac{\mu_u}{2} - \frac{\pi}{4} \lambda v [\lambda_1 (\lambda' v) - \lambda_1 (\lambda' v)] + \frac{1}{2} \lambda v + \frac{\pi \lambda v}{4} [\lambda_0 (\lambda' v) - \lambda_0 (\lambda' v)],$$

(3.30)

where $\lambda = \frac{\mu_2}{\mu_1}$, $\lambda' = \frac{\lambda}{\mu_u}$, $\mu_u = \mathbb{E}(U)$, $\mu_i = \mathbb{E}(Z_i) > 0$ ($l = 1, 2$), $\lambda_i(\cdot)$ is the Struve function $[99, 8.55]$, and $Y_i(\cdot)$ is the Bessel function of the second kind $[99, 8.403]$.

**Proof.** By having the CDF of $\frac{Z_1}{Z_2}$ from Theorem 2, we need to calculate the following integral:

$$F_V(v) = \int_0^\infty F_{\frac{Z_1}{Z_2}} \left( \frac{v}{u} \right) e^{-\frac{u}{\mu_u}} du = \frac{\mu_u}{2} - \frac{1}{2} \int_0^\infty \frac{\mu_1 u}{\sqrt{\mu_1^2 u^2 + \mu_2^2 v^2}} e^{-\frac{u}{\mu_u}} du$$

$$+ \frac{1}{2} \int_0^\infty \frac{\mu_2 v}{\sqrt{\mu_1^2 u^2 + \mu_2^2 v^2}} e^{-\frac{u}{\mu_u}} du. \quad (3.31)$$

Using [99, 3.366.3] and [100, 2.3.5.6], (3.30) can be achieved. \qed

The only remaining task is to derive the CDF of $\gamma_2$, which is done in the next theorem.

**Theorem 3.** For two independent Exponentially distributed random variables $Z_1$ and $Z_2$, and real constant $C$, the CDF of $Z'' = \frac{Z_1 Z_2}{Z_2 + C}$ is:

$$F_{Z''}(z) = 1 - \frac{1}{\mu_2} e^{-\frac{z}{\mu_2}} 2 \left( \frac{\mu_2}{\mu_1} \right) C_1 \left( \frac{1}{\mu_1 \mu_2} Cz \right),$$

(3.32)
where \( \mu_l = E(Z_l) > 0 \) \((l = 1, 2)\) and \( K_1(\cdot) \) is the modified Bessel function of the second kind.

**Proof.** Using the definition of the CDF:

\[
F_{Z''}(z) = \Pr \left( \frac{z_1 z_2}{z_2 + C} < z \right) = \int_0^{\infty} \Pr \left( \frac{z_1 z_2}{z_2 + C} < z | z_2 \right) f_{z_2}(z_2) \, dz_2, \tag{3.33}
\]

and by substituting the exponential distribution, we have:

\[
F_{Z''}(z) = \int_0^{\infty} \frac{1}{\mu_2} \Pr \left( z_1 < z \left( 1 + \frac{C}{z_2} \right) | z_2 \right) e^{-\frac{z_2}{\mu_2}} \, dz_2, \tag{3.34}
\]

which results in:

\[
F_{Z''}(z) = \frac{1}{\mu_2} \int_0^{\infty} e^{-\frac{z_2}{\mu_2}} \, dz_2 - \frac{1}{\mu_2} \int_0^{\infty} \exp \left( -\frac{C}{\mu_1 z_2} - \frac{z_2}{\mu_2} \right) \, dz_2. \tag{3.35}
\]

The first integral in (3.35) can be easily calculated to be \( \frac{1}{2} \). The second integral can be derived using [100, 2.3.16.1], and finally yields (3.32).

It can be easily observed that the random variables \( \gamma_i \) for \( i = 1, 3 \) and 4 are similar to \( V \) in Lemma 7 by assuming small \( c \) and \( c' \). Therefore, \( F_{\gamma_i}(\gamma_i) \) for \( i = 1, 3 \) and 4 follows the CDF in (3.30). By substituting this CDF into (3.18), the average PEPs can be derived. Using integrations \([99, \text{eq 3.361, 6.621.2, 6.823.1}]\) and some manipulations, the average PEPs for \( i = 1, 3 \) and 4 are derived as in (3.36). In this equation, \( Q_p^q(x) \) is the Associated Legendre function of the second kind \([99, 8.7]\), and \( \genfrac{p}{q}{\delta_1, \cdots, \delta_p; \beta_1, \cdots, \beta_q; z} denotes Generalized Hypergeometric Series [99, 9.14]. It is worth mentioning that by considering the PSK modulation scheme, we have \( \lambda = \frac{1}{2} \), and accordingly \( \lambda' = \frac{1}{2\sigma_{\text{SI}}} \).
\[ PEP_i = \frac{a_i \sigma_{\text{SR}}^2 + \lambda}{4a_i} - \frac{5\lambda\lambda'^2}{4} \frac{\sigma_{\text{SR}}^2}{a_i^3} F_2 \left( \frac{1}{4}, \frac{3}{4}; \frac{1}{2}, \frac{1}{2}; -\frac{4\lambda'^2}{a_i^2} \right) \]
\[ - \frac{3\lambda}{16} \sqrt{\frac{a_i}{2}} \left( \lambda'^2 + \frac{a_i^2}{4} \right)^{-\frac{3}{4}} Q_{\frac{1}{2}}^{-1} \left[ \frac{a_i}{2} \left( \frac{a_i^2}{4} + \lambda'^2 \right)^{\frac{1}{2}} \right] \]
\[ + \frac{3\lambda\lambda'}{4a_i^2} \frac{\sigma_{\text{SR}}^2}{a_i^3} F_2 \left( \frac{1}{4}, \frac{3}{4}; \frac{1}{2}, \frac{1}{2}; -\frac{4\lambda'^2}{a_i^2} \right) \]
\[ + \frac{\lambda}{8} \sqrt{\frac{a_i}{2}} \left( \lambda'^2 + \frac{a_i^2}{4} \right)^{-\frac{3}{4}} Q_{\frac{1}{2}}^0 \left[ \frac{a_i}{2} \left( \lambda'^2 + \frac{a_i^2}{4} \right)^{-\frac{1}{2}} \right], \quad i = 1, 3, 4 \]

\[ PEP_{2}(p, p') = \frac{1}{2} - \frac{c}{4\sigma_{\text{RD}}^2} \sqrt{\frac{4a_2(p, p')\sigma_{\text{SR}}^2}{(2 + \sigma_{\text{SR}}^2 a_2(p, p'))^3}} \exp \left( \frac{c}{\sigma_{\text{RD}}^2 (2 + \sigma_{\text{SR}}^2 a_2(p, p'))} \right) \exp \left( \frac{c}{\sigma_{\text{RD}}^2 (2 + \sigma_{\text{SR}}^2 a_2(p, p'))} \right) \]
\[ \times \left[ K_1 \left( \frac{c}{\sigma_{\text{RD}}^2 (2 + \sigma_{\text{SR}}^2 a_2(p, p'))} \right) - K_0 \left( \frac{c}{\sigma_{\text{RD}}^2 (2 + \sigma_{\text{SR}}^2 a_2(p, p'))} \right) \right] \]

For calculating the average PEP, for \( i = 2 \), we can employ \( F_{\gamma_2}(\gamma_2) \) from Theorem 3, and substitute it in (3.18). Then, the resulted integral can be calculated by taking the integral of the Bessel function [99, 6.614.5]. In this equation, \( K_0 \) and \( K_1 \) are modified Bessel functions of the second kind [99, 8.407]. This results in the average PEP equation shown in (3.37).

By having the average PEPs, we can now easily calculate the total ASER at the primary and secondary receivers. The ASER of the primary is obtained using
In (3.38), there are two error terms. The first term is not a function of the APM symbols, while the second term is a function of the transmitted and received APM symbols, \( p \) and \( p' \). Although the instantaneous pair-wise error \( \text{PEP}_1 \) in (3.11) depends on the detected APM symbol, it only affects the correlation coefficient between the random variables in the numerator and denominator inside the Q-function. As the result of Lemma 7, the correlation coefficient for the case of PSK modulation is \( \frac{1}{2} \). Therefore, the effect of the APM symbol in this case can be seen as two different phase rotations on the channels \( h^{(s)}_{\text{RD}} \) and \( h^{(s')}_{\text{RD}} \), which do not impact the mean of the random variables in Theorem 2. Hence, the first term in the average error probability of the primary is independent of the primary APM symbol which is essentially the result of an incorrect channel detection. Moreover, the second term of error at the primary is due to incorrect detection in the APM domain only by considering the Euclidean distance \(|x(p) - x(p')|\) in (3.12). Similarly, the ASER of the secondary is derived by substituting (3.36) into (3.13) as:

\[
\bar{\mathcal{P}}_{\text{Secondary}} \leq (N_R - 1) (M - 1) \text{PEP}_3 + (N_R - 1) \text{PEP}_4.
\] (3.39)

Interestingly, at the secondary, the ASER does not depend on the Euclidean distance of the APM symbols. Since the random variables \( \gamma_1 \) and \( \gamma_3 \) have essentially the same
distributions, the first term of error in (3.39) has the same form as the first term of (3.38). Also, the second term of error only depends on the absolute value of the APM symbol as defined in (22). However, both terms of error depend on the Euclidean distances of the channels. To get a better insight into the performance of the proposed scheme, in the next section, we derive closed-form upperbounds on the ASERs of the primary and the secondary in high SNR regime.

3.4 Asymptotic Analysis of ASER

The total ASERs of the primary and secondary in Section IV contain the Generalized Hypergeometric and Bessel functions. In this section, we analyze the performance of the receivers at high SNRs to derive simple closed-form bounds on the ASERs.

Lemma 8. At high SNRs (large $a_i$), the average $\text{PEP}_i$ for $i = 1, 3$ and $4$ are simplified to:

$$p_{ei} = \frac{\lambda}{4a_i} \left( \ln \left( \frac{4a_i}{\lambda'} \right) - 1 \right).$$ (3.40)

Proof. See Appendix E.

By replacing $\lambda$ and $\lambda'$ for PSK modulation in (3.40), the average PEP for $i = 1, 3$ and $4$ becomes:

$$p_{ei} = \frac{\sigma^2_R}{4P_S} \left( \ln \left( \frac{4P_S}{\sigma^2_R} \right) - 1 \right).$$ (3.41)

Lemma 9. At high SNRs (large $a_2(p, p')$), the average $\text{PEP}_2$ is simplified to:

$$p_{e2(p, p')} = \frac{1}{a_2(p, p')} \frac{c}{2\sigma^2_{SR}\sigma^2_{RD}} \left( \ln \left( \frac{a_2(p, p')2\sigma^2_{SR}\sigma^2_{RD}}{c} \right) - \gamma - 1 \right),$$ (3.42)

where $\gamma$ is the Euler-Mascheroni constant [99, 9.63].

74
Table 3.1: Maximum number of antennas at the secondary transmitter for different primary APM sizes and error thresholds, unit channel variances and $P_t/N_0 = 50$ dB.

<table>
<thead>
<tr>
<th>$M$</th>
<th>$N_R$</th>
<th>$N_R$</th>
<th>$N_R$</th>
<th>$N_R$</th>
<th>$N/A$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>17</td>
<td>9</td>
<td>1</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>17</td>
<td>9</td>
<td>5</td>
<td>3</td>
<td>N/A</td>
</tr>
<tr>
<td>8</td>
<td>34</td>
<td>11</td>
<td>4</td>
<td>3</td>
<td>N/A</td>
</tr>
<tr>
<td>16</td>
<td>11</td>
<td>4</td>
<td>1</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>11</td>
<td>4</td>
<td>1</td>
<td>N/A</td>
<td></td>
</tr>
</tbody>
</table>

Proof. For high SNR regime, the $\text{PEP}_2(p, p')$ in (3.37) can be written as:

$$p_{e2}(p, p') = \frac{1}{2} - \frac{1}{2} a_2(p, p') \frac{c}{\sigma^2_{SR} \sigma^2_{RD}} \exp \left( \frac{1}{a_2(p, p')} \frac{c}{\sigma^2_{SR} \sigma^2_{RD}} \right) \left( K_1 \left( \frac{1}{a_2(p, p')} \frac{c}{\sigma^2_{SR} \sigma^2_{RD}} \right) - K_0 \left( \frac{1}{a_2(p, p')} \frac{c}{\sigma^2_{SR} \sigma^2_{RD}} \right) \right).$$

(3.43)

The Generalized Puiseux series [99, 8.446] for a small $r$ implies that:

$$\omega r e^{\omega r} [K_1 (\omega r) - K_0 (\omega r)] \approx 1 + \omega r \left( \ln \left( \frac{\omega}{2} \right) + \ln (r) + \gamma + 1 \right).$$

(3.44)

Therefore, using the Generalized Puiseux series in (3.43), and substituting $\omega = \frac{c}{\sigma^2_{SR} \sigma^2_{RD}}$ and $r = \frac{1}{a_2(p, p')}$ yields (3.42).

As mentioned, $\text{PEP}_2$ at the primary receiver is a function of the distances between the APM symbols. To drive an upperbound on the asymptotic ASER, we consider the minimum distance of PSK symbols; i.e. we consider:

$$a_{2, \min} = \frac{P_S \sin \left( \frac{\pi}{M} \right)}{\sigma^2_R}.$$

(3.45)

The ASER of the primary at high SNRs can be upperbounded as:

$$\overline{P_{\text{Primary}}} \leq (N_R - 1) (M - 1) p_{e1} + (M - 1) p_{e2},$$

(3.46)
where $p_{e2}$ is the function of the modulation size. Considering the minimum distance between the APM symbols, $p_{e2}$ can be written as:

$$p_{e2} = \frac{c\sigma^2_R}{2P_S \sin \left( \frac{\pi}{M} \right) \sigma^2_{SR}\sigma^2_{RD}} \times \left[ \ln \left( \frac{2P_S \sin \left( \frac{\pi}{M} \right) \sigma^2_{SR}\sigma^2_{RD}}{c\sigma^2_R} \right) - \gamma - 1 \right]. \quad (3.47)$$

The asymptotic ASER at the secondary receiver can be upperbounded using Lemma 8 as:

$$\bar{P}_{\text{Secondary}} \leq (N_R - 1)(M - 1)p_{e3} + (N_R - 1)p_{e4}. \quad (3.48)$$

Since we consider the PSK modulation at the primary transmitter, we have $a_3 = a_4$ and so, at the secondary receiver $p_{e3} = p_{e4}$. The upperbound on the asymptotic ASER at the secondary receiver can then be simplified to:

$$\bar{P}_{\text{Secondary}} \leq (N_R - 1)M p_{e3}. \quad (3.49)$$

### 3.5 System Design

It is possible to design critical system parameters at primary and secondary to achieve the maximum possible transmission rates while guaranteeing an error performance target for the primary and possibly the secondary networks. For a given primary constellation size $M$, the secondary transmitter can adjust its spatial domain constellation size, i.e., $N_R$, to satisfy primary performance target while achieving the maximum possible transmission rate in the secondary system. In other words, maximum $N_R$ for a fixed APM size $M$, can be selected such that the probability of error at the primary receiver is guaranteed to be less than $\bar{P}_P$. On the other hand, the secondary system must have a reasonable error performance, i.e. $\bar{P}_S$, to be able to
transmit its own information via spatial domain. It can be easily shown from (3.46) and (3.49) that the maximum number of antennas of the secondary transmitter should follow:

\[
N_R = \left\lfloor \min\left\{ \frac{\hat{P}_P - (M - 1) M p_{e2}}{(M - 1) p_{e1}}, \frac{\hat{P}_S}{M p_{e3}} \right\} \right\rfloor + 1, \tag{3.50}
\]

where \(\lfloor \cdot \rfloor\) denotes the floor function. For example, the maximum possible number of antennas from (3.50) for some scenarios is provided in Table. 3.1. \(P_t\) in the table refers to total transmit power, i.e. \(P_t = P_S + P_R\). As seen, for small APM sizes, the secondary network is able to transmit its own information at a higher rate, while helping the primary network by retransmitting its APM symbols. For some larger APM sizes, on the other hand, there are cases where \(N_R \leq 1\). In such cases, the secondary cannot achieve a meaningful transmission rate without compromising the error performance target of the primary. In such cases, it might be possible for primary and secondary to enter into a negotiations such that by picking the critical system parameters, such as target performance and transmission power, both can achieve acceptable transmission rates.

Another possible design scenario is when the primary is aware of the system parameter of the secondary. In such a case the primary transmitter is able to select its best APM constellation size, \(M\), to achieve a certain performance guarantee for a fixed number of secondary transmit antennas, \(N_R\). For example, Fig. 3.2 shows \(M\) as a function of \(N_R\) and the SER of the primary. This figure shows the effect of increase in possible primary APM size as an increasing function of \(N_R\) and its SER.
3.6 Simulation Results

In this section, we numerically evaluate the SER of the primary and secondary systems using the Monte Carlo simulation, and compare the results with the exact analytical derivations of Section IV, as well as the asymptotic analysis of Section V. Throughout the simulations, we assume equi-power noises at all receivers, i.e., $\sigma_R^2 = \sigma_D^2 = \sigma_C^2 = N_0$. We also consider that $P_t = P_S + P_R$, and plot the SER versus $P_t/N_0$. It is worth mentioning that the degree and order of the Associated Legendre functions of the second kind in the formulation of the PEPs in (3.36) are not integer. For the ease of simulations using MATLAB, we convert the Associated Legendre functions of the
second kind into Hypergeometric form [101, eq 59] as follows:

\[ Q^{-1}_{\frac{1}{2}}(r) = \frac{1}{4} \sqrt{\frac{\pi}{2}} \frac{\Gamma\left(\frac{1}{4}\right)}{\Gamma\left(\frac{3}{4}\right)} (1 - r^2)^{-\frac{1}{2}} F\left(\frac{1}{4}, -\frac{3}{4}; \frac{1}{2}; r^2\right) \]
\[ + \frac{1}{2} \sqrt{\frac{\pi}{2}} \frac{\Gamma\left(\frac{3}{4}\right)}{\Gamma\left(\frac{5}{4}\right)} (1 - r^2)^{-\frac{1}{2}} r F\left(-\frac{1}{4}, \frac{3}{4}; \frac{3}{2}; r^2\right), \]

where \( \Gamma(\cdot) \) is gamma function [99, 6.4] and for the zero order Associated Legendre functions of the second kind, we have:

\[ Q^0_{\frac{1}{2}}(r) = -\frac{1}{2} \sqrt{\frac{\pi}{2}} \frac{\Gamma\left(\frac{3}{4}\right)}{\Gamma\left(\frac{5}{4}\right)} F\left(\frac{3}{4}, -\frac{1}{4}; \frac{1}{2}; r^2\right) \]
\[ + \sqrt{\frac{\pi}{2}} \frac{\Gamma\left(\frac{5}{4}\right)}{\Gamma\left(\frac{3}{4}\right)} r F\left(-\frac{1}{4}, \frac{5}{4}; \frac{3}{2}; r^2\right). \]

Figs. 3.3 and 3.4 show the ASER of the primary and secondary systems. The first observation from these figures is that the asymptotic results closely follow the analytical closed-form derivations in about all the range of SNRs. This result shows that the approximations in (44), (66) and (67) are tight in practical SNR ranges. However, at low SNRs, especially when the size of both the APM and spatial domains are large, the analytical results do not completely match the numerical results. This is because the union bound is not tight at low SNRs. Also, since we consider small \( c \) and \( c' \) in deriving some of the PEP expressions, the analytical curves are shifted slightly below the numerically derived curves.

Figure 3.3 shows the performance of the primary network for different sizes of spatial and APM domains. The figure shows that considering a fixed PSK constellation size \( M \), for transmitting every additional information bit of the secondary (i.e. doubling the number of transmit antennas of the secondary), the total required
transmission power to maintain the performance of the primary network, $P_t$, increases by 4dB. The ASER performance of the secondary system for different sizes of spatial APM domains is depicted in Fig. 3.4. It can be seen from this figure that for a fixed spatial domain size (i.e. fixed number of secondary transmit antennas), for transmitting every additional information bit of the primary information (i.e. doubling the APM size) requires about 3dB additional total transmission power to maintain the same SER at the secondary receiver. Also, it can be seen that the slope of the curves are the same in this figure, and this implies that the achieved diversity is fixed for different transmit antenna sizes.

The effect of the number of secondary antennas on the ASER of the primary and the secondary systems is illustrated in Fig. 3.5 at a fixed $P_t/N_0$ of 30dB. This figure
Figure 3.4: ASERs of the secondary system for different sizes of APM and spatial domains.

can be used to pick the proper number of secondary transmit antennas as discussed in Section VI. For instance, if we target a maximum SER of $10^{-3}$ at both receivers, assuming $M = 2$, the only possible number of secondary antennas is $N_R = 2$, which keeps both the SERs below the threshold. This figure also shows that the proper number of secondary antennas is very sensitive to the target SER threshold, and consequently, the transmission power, such that by reducing the threshold to $0.9 \times 10^{-3}$ and assuming the same $M = 2$, we can now accommodate $N_R = 4$, which doubles the transmission rate of the secondary. This figure also illustrates the performance of the primary receiver in the extreme case of $N_R = 1$, i.e. when the secondary transmitter only works like a simple relay for the primary and does not transmit its own informa-
Figure 3.5: ASERs of the primary and secondary systems versus number of secondary antennas ($N_R$) for different APM sizes; $P_t/N_0 = 30$dB.

Comparing this case, which can be considered as a simple primary AF relaying, with the proposed overlay CR scheme ($N_R > 1$) shows that increasing the number of antennas at the secondary transmitter not only provides a healthy transmission rate for the secondary but also improves the SER performance of the primary.

Even though, we derived the closed-form ASER for the case of single-antenna receivers, the proposed scheme is equally applicable when the receivers are equipped with multiple antenna. Although deriving the CDF of the random variables $\gamma_i$ ($i = 1, \cdots, 4$) in (18)-(21) for the case of multiple receive antennas is cumbersome, the detector and the instantaneous error probabilities in this work have been derived for the general case of multi-antenna receivers. In Fig. 3.6, we illustrate the error
performance numerically for the case of multiple receive antennas. It is worth mentioning that considering multiple receive antennas and using the joint ML detection provides a diversity gain as shown in Fig. 3.6. It has been shown in [39, Prop. 4] that the diversity is $\frac{m_{\text{Nak}} N_D}{\log_2 (M)}$, where $m_{\text{Nak}}$ is Nakagami fading parameter, and for Rayleigh fading $m_{\text{Nak}} = 1$.

For the case of multiple antennas at the secondary receiver, the secondary receiver can only detect its corresponding transmitter’s antenna index and the detection of the APM symbol may not be necessary. For the case of more than one receive antenna, approaches such as maximum ratio combining (MRC) [102, 103, 104, 105] can be used to detect the index of the transmitting antenna without the need for detecting the APM symbol. This detector is used to estimate both the transmitting antenna index and the APM symbol in two stages. The antenna index can be estimated by:

$$\tilde{s} = \arg \max_s \left| \frac{\mathbf{h}_{\text{RC}}^{(s)H} \mathbf{y}_C}{\|\mathbf{h}_{\text{RC}}^{(s)}\|} \right|. \quad (3.53)$$

This detector only estimates the antenna index of the relay node in each symbol interval and does not require the primary information for detection. As seen in Fig. 3.6, while it is a sub-optimal detector, MRC detector provides a reasonable performance compared to optimal joint ML detector which has been also pointed out in [105] and [104]. It has been shown in these references that the diversity order of the MRC detection is always one less than that of the ML detection which can be also observed from this figure.
To better evaluate the performance of the proposed overlay CR scheme, we compare its performance with that of an underlay CR scheme, in which the primary and the secondary networks are working separately in the same spectrum band. For the overlay scheme, we assume $N_R = 4$, and for the considered underlay scenario, we assume that the primary network has its own relay node and the secondary network consists of a pair of transmitter and receiver. It is assumed that all the nodes are equipped with single antenna and use Quadrature PSK (i.e. $M = 4$) modulation. To have equal transmission rates at the primary and secondary networks, we also assume that the secondary is only transmitting in the second phase of primary relaying. As can be seen in Fig. 3.7, the performance of the primary network in the case of the proposed overlay scheme is about 15dB better than its performance in the case of the underlay scenario. The performance degradation in the latter case is due to the
Figure 3.7: Performance comparison of the proposed overlay scheme and an underlay CR scheme, $N_R = M = 4$.

interference of the secondary in the second phase. However, if we consider the case where the secondary is not present, the performance of the primary receiver can be 10dB better than its performance in the overlay scenario. This is expected because in the proposed overlay approach the primary receiver should detect both spatial and APM domains and thereby incur more errors. On the other hand, the proposed overlay scheme provides a significantly better secondary performance compared to the underlay approach. In fact, the performance of the underlay scheme is limited by the interference from the primary network especially at high transmission powers. However, in the proposed overlay approach, the secondary information is transmitted in the spatial domain, and its performance is not compromised by the primary interference.
CHAPTER IV
COGNITIVE MOLECULAR COMMUNICATIONS

In this chapter, inspired by the spectrum sensing in wireless communications, we provide a cognitive molecular system in which the secondary nanonetwork measures the number of molecules as a criterion to decide about the presence or absence of the primary nanonetworks using molecular energy detection. We firstly discuss the importance and the feasibility of utilizing cognition for the future of molecular communications and also provide the system model of the proposed scheme. We then divide the sensing mechanism based on the availability of timing information at the sensing nanodevice into two cases: synchronous and asynchronous molecular sensing. In the synchronous sensing, we provide two mechanisms of sensing by considering the sensing window time smaller or larger than the primary pulse-width, i.e. single-pulse and multiple-pulse sensing, respectively. We also derive a likelihood ratio test based on the Bayesian approach, when the timing information of the primary is not available. The performance of the proposed sensing schemes in a molecular environment is then evaluated using numerical simulations.
4.1 System Model and Problem Formulation

In this section, we introduce the concept of cognition in the nanoscale devices and some practical models are provided. This proposed model is then described by mathematical formulation of the concentration-based transmission in molecular communications.

4.1.1 Nanoscale Cognition

In nature, chemical signals are used for data transmission at macroscopic and microscopic scales for inter-cellular or intra-cellular communications. In most cases, signaling cells emit signal molecules into the extracellular medium (see e.g. [106, Sec. 1.2]). The emitted molecules freely diffuse in the medium and may reach and act on distant target cells. Such a signaling is called as paracrine signaling as shown in Fig. 4.1. The emitted signals may act as local mediators which interact only with cells in the surrounding environment of the signaling cell. In this case, the signaling is referred to as autocrine signaling. For example, cancer cells often follow the autocrine signaling to stimulate their own survival and proliferation. On the other hand, the inter-cellular communications in nature can be performed by endocrine signaling, in which endocrine (signaling) cells release hormone (signaling) molecules into the bloodstream as shown in Fig. 4.2. This enables the hormone molecules to reach to and interact with distant target cells.

Recently, these types of inter-cellular communications have been the source of inspiration for researchers to design information exchange schemes for nanomachines.
One of the first works on engineering an experimental demonstration of molecular communication at microscales is reported in [107]. Considering demands for synthetic biology and its applications, such as biological computation, a simple inter-cellular communication has been provided for multicellular computation [108, 109]. In [110], artificial cells are used as relays (translates) to send a message to Escherichia coli (E. coli) cells. In this scheme, the intended message is first detected and decoded by the artificial cells, and then relayed to the E. coli cells. A comprehensive survey on recent applications of molecular communications can be found in [69]. Besides cell-to-cell communications, there has been a significant interest in using molecular communications as a means for interaction among nanomachines [60], for application in tissue engineering, targeted drug delivery, enhanced immune systems and micro-electromechanical systems (MEMS).

Given that in several applications, the molecular communications should occur inside a biological environment, such as the human body, choosing an appropriate molecule for data transmission with minimum invasiveness to biological cells is an important factor. On the other hand, the types of molecules that can be used for signaling in a biological environment may be very limited. Researchers mostly get inspired by the signaling mechanisms of inter-cellular communications and select molecules which are bio-compatible with no, or minimum, risk of harm. This can significantly limits the number of independent nanonetworks that can coexist in the same environment as each network should use its unique type of signaling molecule.

As a practical scenario, consider a drug delivery system, which consists of
Figure 4.1: A pair of cognitive nanomachines in the vicinity of a primary nanonetwork with paracrine signaling.

nanoscale actuators, monitoring/control nanomachines and target cells (see e.g. [63] and the references therein). In this system, the monitoring node is able to measure the effects of drugs on the target by measuring specific quantities and to control the emission rate. We assume that the information exchange between monitoring node and actuator is carried by molecules. On the other hand, consider that another nanonetwork for a different purpose, e.g. for tissue engineering, is required in the same body which uses the same type of molecule for its interconnection. For instance, the bone regeneration is considered in order to accelerate the regeneration process and to control such a process in case of incomplete regeneration [111]. The release of
stimulants in this system is controlled by nanomachines connected through molecular communications.

In such a scenario, inserting the nanonetwork for tissue engineering can be challenging as it can cause significant interference to the drug delivery system. However, if the second nanonetwork has a cognition ability to only transmit when the first nanonetwork is idle, then the two networks can coexist in the same body. In this work, we consider adding the cognition capability to the secondary nanonetwork, i.e. the tissue engineering nanomachines in the above example, to adapt its communications based on the sensing of the molecular environment and to use the channel when the primary, i.e. the drug delivery transmission in the above example, is not present. One can establish an analogy between cognition in molecular communications and the cognitive radio communications, which provides an effective tool to solve the problem of spectrum crunch [7, 6]. This concept can be intelligently utilized in molecular communications, where there is a limitation on the number of molecules for carrying information. In this study, we propose methods to sense the presence of primary nanonetworks by monitoring the concentration of molecules at the secondary transmitter. The secondary nanonetwork can opportunistically use the channel for its transmission when primary is not transmitting.

4.1.2 Problem Formulation

In this work, we consider a primary nanonetwork as the network that has been installed a priori or otherwise possesses a higher priority, that operates under the
paracrine or endocrine signaling as shown in Fig. 4.1 and Fig. 4.2, respectively. A pair of the secondary or cognitive transmitter and receiver operate in the vicinity of the primary nanonetwork. We assume that both networks use the same type of information carrying molecules.

**Primary Nanonetwork**

We consider a simple pulse-based modulation scheme for the primary nanonetwork, in which the transmitter releases molecules at the beginning of each time period $T_p$. This can be considered as a spike in molecular concentration at the transmitter which propagates throughout the medium. We assume that the transmitter can perfectly control the number of emitted molecules. Considering $r$ as the distance between the transmitter and receiver, it is assumed that the dimension of the receiver along the
$x$-axis is very small and in the order of $r/100$. The molecules are propagated through the medium with the drift velocity of $v$ from transmitter to receiver. The non-zero drift velocity can be practical for the endocrine scenario (Fig. 4.2), in which the medium flow affects the propagation of molecules. Diffusion coefficient is defined by $D = K_B T_a / 6\pi \eta d_m$, where $K_B$ is Boltzmann constant, $T_a$ is the absolute temperature, $\eta$ is the velocity constant of the fluid medium, and $d_m$ is the radius of a molecule. The propagation of molecules follows the Brownian motion, which affects the random movement of molecules based on the diffusion coefficient of the medium. This effect is modeled as a Wiener process with variance $\sigma^2 = 2D$ [112, 113].

Assuming that the receiver is perfectly absorbing, the arrival time of a molecule follows an inverse Gaussian distribution with the probability distribution function (PDF) given by:

$$f(t|\mu, \lambda) = \sqrt{\lambda/2\pi t^3} e^{-\lambda(t-\mu)^2/2\mu^2}, \quad t > 0,$$

where $\mu = \frac{r}{v}$ and $\lambda = \frac{r^2}{\sigma^2}$. It is worth mentioning that when the velocity is zero ($v = 0$), the distribution of travelling time of molecule is no longer inverse Gaussian and it can be simply written as [113]:

$$f(t|r, D) = \frac{r}{\sqrt{4\pi D t^3}} e^{-\frac{r^2}{4Dt}},$$

In this study, we consider the concentration-based modulation scheme at the primary nanonetwork. Sometimes this modulation scheme is called concentration shift-keying (CSK) modulation. In this scheme, the concentration, or equivalently, the number of the received molecules identifies the transmitted symbol [64, 66, 67, 68].
The receiver measures the signal energy in each pulse, and counts the number of molecules in each time interval. The probability of arrival of a single molecule in the first pulse time \([0, T_p]\) can be calculated by:

\[
p(0, T_p) = \int_0^{T_p} f(t)dt.
\]

The probability of arrival in the first pulse time can be derived using the cumulative distribution function (CDF) of an inverse Gaussian random variable as \(p(0, T_p) = F(T_p) - F(0)\), where the CDF of the random variable \(t\) is given by:

\[
F(t|\mu, \lambda) = \frac{1}{2} \left[ 1 + \text{erf} \left( \sqrt{\frac{\lambda}{2t}} \left( \frac{t}{\mu} - 1 \right) \right) \right] + \frac{1}{2} e^{2\lambda/\mu} \left[ 1 - \text{erf} \left( \sqrt{\frac{\lambda}{2t}} \left( \frac{t}{\mu} + 1 \right) \right) \right],
\]

and error function is defined as \(\text{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-x^2} dx\). For the case of zero velocity, this can be easily written as:

\[
p(0, T_p) = \int_0^{T_p} \frac{r}{\sqrt{4\pi DT_p^3}} e^{-\frac{r^2}{4DT_p}} dt = \text{erfc} \left( \sqrt{\frac{r^2}{4DT_p}} \right),
\]

where the complementary error function is defined as \(\text{erfc}(x) = 1 - \text{erf}(x)\). For the sake of simplicity, in the rest of this chapter, we consider the probability of arrival as a function of time interval \([a, b]\) as \(p(a, b) = \int_a^b f(t)dt = F(b) - F(a)\).

For a better understanding of the \(p(\cdot, \cdot)\) function, we plot this probability for different window sizes \(\alpha T_p \ (\alpha \in (0, 1))\) over two consecutive time periods versus the drift velocity of medium in Fig. 4.3. As shown, the probability of arrival of molecules is maximized in the first time period, when the direction of the drift velocity is positive, and hence, it increases the arrival probability (or equivalently energy) at the
Figure 4.3: The probability of arrival of one molecule at the distance of \( r_s = 30 \, \mu \text{m} \) versus velocity of medium, \( D = 1 \, \text{nm}^2/\text{ns} \).

beginning of the first time period. This also results in a reduction in the probability of arrival in the second time period.

We consider that the primary uses an \( M \)-level pulse modulation scheme, in which \( \log_2(M) \) bits are transmitted depending on the number of released molecules (i.e. trials). The distribution of the number of received molecules (energy of the signal) in time interval \( (0, T_p) \) can be described by a binomial distribution, \( B(Q_i, p(0, T_p)) \), where \( Q_i, i = 0, 1, \ldots, M - 1 \) (\( Q_0 < Q_1 < \cdots < Q_{M-1} \)), is the number of the trials when the \( i^{th} \) symbol is transmitted, and \( p(0, T_p) \) is the success probability in each trial [114]. It is well known that for large \( Q_i \)'s, the binomial distribution can be approximated with a normal distribution as \( \mathcal{N}(p(0, T_p)Q_i, p(0, T_p)(1 - p(0, T_p))Q_i) \).
The detection scheme at the primary receiver can be designed based on the maximum concentration point or the total received energy in each pulse interval. The detection can be done by comparing the received energy with predefined thresholds \cite{68}. If a total of \( Q \) molecules at time instance \( t = 0 \) are transmitted, the received concentration is initially zero and quickly increases until reaching a maximum. The molecular concentration then slowly decreases over time due to the effect of diffusion \cite{67}. Interestingly, the concentration at the maximum point is only proportional to the number of released molecules \( Q \) and is not a function of the diffusion coefficient \( D \) \cite{67}. On the other hand, it can be easily shown that the time of the maximum concentration is \( t_{max} = \frac{r^2}{6D} \), which implies that the pulse-width decreases in a medium with higher diffusion coefficient.

Secondary Nanonetwork

We assume that the secondary nanonetwork consists of a pair of transmitter and receiver as shown in Figs. 4.1 and 4.2. The secondary transmitter is equipped with a sensing nanodevice to measure the energy of the received signal. Based on the measured energy, it decides to use, or not to use, the channel for communicating with its corresponding receiver. The amount of released molecules by the secondary transmitter is also under control and it is always selected based on the maximum interference that the primary receiver can endure. The modulation scheme of the secondary is arbitrary, and considering the constraint on the number of released molecules at the transmitter, both concentration-based and timing-based \cite{64, 65, 66, 95}.
modulation schemes can be applied. In the case of CSK modulation with zero medium velocity, the secondary nanonetwork requires to maintain its maximum received energy at the primary receiver below a threshold. If the maximum released concentration of the secondary is \( Q_{CR}^{M-1} \) for an \( M \)-level CSK modulation scheme, and considering the threshold of \( \gamma \) as the maximum interference allowed at the primary receiver, the number of released molecules at the secondary transmitter is bounded as:

\[
Q_{CR}^{M-1} \leq \frac{\gamma}{\text{erfc} \left( \sqrt{\frac{r^2}{4DT_p}} \right)}.
\]

(4.6)

Note that in writing (4.6), it is assumed that the pulse width of the secondary is also \( T_p \). The equality holds when the primary receiver receives synchronously with negligible inter-symbol interference (ISI). Since the mechanism of detection in the primary receiver is based on the pulse energy, for the case of timing-based modulation, we can also control the number of released molecules at the secondary transmitter to meet the interference requirement of the primary.

### 4.2 Synchronous Molecular Channel Sensing

In this section, we consider the energy detection for sensing the molecular channel, when the primary timing information is available at the sensing nanodevice. The performance of the sensing scheme affects the transmission of the secondary network; however, since the interference to the primary is always kept below a threshold, the performance of the primary is not affected significantly. Based on the prior knowledge
Figure 4.4: The sensing timing scheme for three different methods.

of the primary parameters at the secondary sensing nanodevice, we propose three sensing strategies.

4.2.1 Synchronous Single-Pulse Sensing

If we assume that the sensing nanodevice is able to synchronize the energy detection with the primary pulses, sensing can be applied within a portion of the primary
transmission interval, i.e. \( \alpha T_p, 0 \leq \alpha \leq 1 \), to perform a hypothesis test. The secondary transmitter can then use the remaining part of the transmission interval for its own transmission, when the primary transmission is not present. The detection problem then becomes a hypothesis test, where hypotheses \( H_0 \) and \( H_1 \) represent the cases of inactive and active primary, respectively. In the former case, i.e. for \( H_0 \), we consider the worst case scenario when the previous primary pulse is active. The sensing window then becomes \( t \in [T_p, (\alpha + 1)T_p] \). Therefore, the hypothesis test can be applied on the energy of the received signal \( e_s \) at the secondary sensing nanodevice to identify the status of channel as follows:

\[
H_0: e_s \sim \mathcal{N}(\bar{\mu}_i, \bar{\sigma}_i^2), i = 0, 1, \ldots, M - 1,
\]

\[
H_1: e_s \sim \mathcal{N}(\mu_i, \sigma_i^2), i = 0, 1, \ldots, M - 1,
\]

where \( \bar{\mu}_i = p(T_p, (\alpha + 1)T_p) Q_i \), \( \bar{\sigma}_i^2 = p(T_p, (\alpha + 1)T_p) [1 - p(T_p, (\alpha + 1)T_p)] Q_i \), \( \mu_i = p(0, \alpha T_p) Q_i \) and \( \sigma_i^2 = p(0, \alpha T_p) [1 - p(0, \alpha T_p)] Q_i \). Assuming that the secondary sensing nanodevice has the a priori knowledge of primary probabilities, the PDF of the received energy given \( H_0 \) for an \( M \)-ary transmission scheme can be written as:

\[
f(e_s|H_0) = \sum_{i=0}^{M-1} f(e_s|i|H_0) = \sum_{i=0}^{M-1} f(e_s|i, H_0) p_{s_i},
\]

where \( p_{s_i}, i = 0, \ldots, M - 1 \) are the a priori probabilities of primary symbols. Considering that \( f(e_s|i, H_0) \) is \( \mathcal{N}(\bar{\mu}_i, \bar{\sigma}_i^2) \), \( e_s \) in \( H_0 \) has a mixture-Gaussian distribution, i.e.:

\[
f(e_s|H_0) = \sum_{i=0}^{M-1} \frac{p_{s_i}}{\sqrt{2\pi \bar{\sigma}_i}} e^{-\frac{(e_s - \bar{\mu}_i)^2}{2\bar{\sigma}_i^2}},
\]
and similarly:

\[ f(e_s|H_1) = \sum_{i=0}^{M-1} \frac{p_{s_i}}{\sqrt{2\pi} \sigma_i} e^{-\frac{(e_s - \mu_i)^2}{2\sigma_i^2}}. \]  

(4.10)

A simple energy detection scheme at the secondary sensing nanodevice can be firstly considered which compares the energy of the received molecules in each sensing interval with a threshold \( \lambda \). In this case, the probability of false alarm, i.e. the probability that the channel is not occupied but mistakenly detected as occupied, can be written based on the calculated PDF of energy as:

\[ P_F = \Pr(e_s > \lambda|H_0) = \sum_{i=0}^{M-1} p_{s_i} Q\left(\frac{\lambda - \bar{\mu}_i}{\bar{\sigma}_i}\right), \]  

(4.11)

where \( Q(x) \) denotes the Q-function defined as \( Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-\frac{t^2}{2}} dt \). Therefore, the probability of false alarm is the summation of Q-functions. For the simple case of binary modulation, it can be written as:

\[ P_F = p_{s_0} Q\left(\frac{\lambda - \bar{\mu}_0}{\bar{\sigma}_0}\right) + p_{s_1} Q\left(\frac{\lambda - \bar{\mu}_1}{\bar{\sigma}_1}\right). \]  

(4.12)

It is worth mentioning that the probability of false alarm is calculated based on the assumption that the previous primary symbol is always present. Therefore, the probability of false alarm in (4.12) is indeed an upperbound. When the probability of primary transmission is known at sensing nanodevice, (4.9) and (4.10) can be adjusted by multiplying them with this probability to have a more accurate sensing scheme.

The probability of correct detection, i.e. the case where the channel is occupied and is correctly detected by the secondary sensing nanodevice, can also be
written as:

\[
P_D = \Pr(e_s > \lambda | \mathcal{H}_1) = \sum_{i=0}^{M-1} p_{s_i} Q\left(\frac{\lambda - \mu_i}{\sigma_i}\right).
\]  

(4.13)

For the case of binary modulation scheme, the probability of detection is:

\[
P_D = \Pr(e_s > \lambda | \mathcal{H}_1) = p_{s_0} Q\left(\frac{\lambda - \mu_0}{\sigma_0}\right) + p_{s_1} Q\left(\frac{\lambda - \mu_1}{\sigma_1}\right).
\]  

(4.14)

Since both probabilities include summation of Q-functions, it is not analytically possible to find the threshold value \(\lambda\) for a certain probability of false alarm \(P_F\). This is different from energy detection in radio communications, where the energy of the received signal typically follows \(\chi^2\)-squared distribution and, for large number of samples, can be approximated as a Gaussian distribution [6]. In that case, the detection threshold can be conveniently calculated for a target \(P_F\).

On the other hand, since the PDF of the received signal in both hypotheses are known, we can use likelihood ratio test (LRT) as an optimal detection scheme in the sensing nanodevice. Based on Neyman–Pearson (NP) theorem [115], the probability of detection is maximized by defining the test criterion for the threshold value \(\gamma\) as:

\[
L(e_s) = \frac{f(e_s | \mathcal{H}_1)}{f(e_s | \mathcal{H}_0)} \overset{\mathcal{H}_1}{\underset{\mathcal{H}_0}{\gtrsim}} \gamma,
\]  

(4.15)

where threshold \(\gamma\) is found from:

\[
P_F = \Pr(L(e_s) > \gamma | \mathcal{H}_0) = \delta,
\]  

(4.16)

where \(\delta\) is a constant. This detection scheme provides the maximum probability of detection for a certain probability of false alarm, and by substituting the PDF of
energy in $H_0$ and $H_1$, we have:

$$L(e_s) = \sum_{i=0}^{M-1} \frac{p_{s_i}}{\sqrt{2\pi}\sigma_i} e^{\frac{(e_s-\mu_i)^2}{2\sigma_i^2}}.$$  

Therefore, for the LRT detection scheme, the detector derives the test criterion by calculating the received energy $e_s$ in each sensing interval and substituting into (4.17). When the probability of channel usage $p_u$ is available at the sensing nanodevice, the threshold can be selected so that the probability of error is minimized [115].

The probability of sensing error is defined as:

$$P_e = P_F \Pr(H_0) + P_{MD} \Pr(H_1),$$  

where $P_{MD}$ is the probability of missed detection and is defined as $P_{MD} = 1 - P_D$.

The optimal sensing is achieved when

$$\gamma = \frac{\Pr(H_0)}{\Pr(H_1)} = \frac{1-p_u}{p_u}.$$  

Example: As shown in Fig. 4.4(c), the synchronous single-pulse sensing is applied in three time-slots, i.e. $S_1$, $S_3$ and $S_4$. If the primary channel occupancy in (b) is considered, the secondary nanonetwork is able to use the channel in $S_4$ for $(1-\alpha)T_p$ time duration in which the primary transmission is not present. Moreover, since the availability of the channel can be sensed in each time-slot, there is no disruption, i.e. interference, to the secondary transmission from the primary as long as there is no missed detection. So, this method can be applied even if the probability of usage of channel by primary is large.
4.2.2 Synchronous Multiple-Pulse Sensing

To achieve a more robust sensing scheme, the sensing can be applied to multiple primary pulses. As shown in Fig. 4.4(d), in this scenario, the secondary sensing nanodevice measures the energy of \( m \) pulses while considering the timing information of the primary is available at the sensing nanodevice. The absence of primary transmitter (\( \mathcal{H}_0 \)) can be considered when there is no transmitted data during \( m \) pulse duration. Accordingly, the hypothesis \( \mathcal{H}_1 \) is chosen when one or more pulse durations (out of \( m \) pulse durations) are occupied by the primary.

Considering that the energy at the \( i \)th pulse is \( e_s(i) \), the total energy of \( m \) pulses can written as:

\[
\bar{e}_s = \sum_{j=1}^{m} e_s(j).
\]  

(4.19)

For the case of hypothesis \( \mathcal{H}_0 \), we consider the case that the channel before the sensing window is occupied by the primary and is empty for the next \( m \) pulses. Therefore, the signal energy follows a Gaussian distribution as \( \bar{e}_s \sim \mathcal{N}(\tilde{\mu}_i, \tilde{\sigma}_i^2) \), where the mean and variance are:

\[
\tilde{\mu}_i = p(T_p, (m + 1)T_p)Q_i,
\]  

(4.20)

\[
\tilde{\sigma}_i^2 = p(T_p, (m + 1)T_p) [1 - p(T_p, (m + 1)T_p)] Q_i.
\]  

Then, the PDF of signal energy in hypothesis \( \mathcal{H}_0 \) is a mixture of Gaussians; i.e.:

\[
f (\bar{e}_s|\mathcal{H}_0) = \sum_{i=0}^{M-1} \frac{p_{si}}{\sqrt{2 \pi \tilde{\sigma}_i^2}} e^{-\frac{(e_s - \tilde{\mu}_i)^2}{2 \tilde{\sigma}_i^2}}.
\]  

(4.21)

For the case of hypothesis \( \mathcal{H}_1 \), we assume that \( n > 0 \) active pulses are present and the probability of channel usage by the primary is \( p_u \). For the case of binary mod-
ulation scheme, the PDF of the energy given \( n \) in \( \mathcal{H}_1 \) can be written as a combination of Gaussian distributions as follows:

\[
f(\bar{e}_s|\mathcal{H}_1, n) = \sum_{r=0}^{n} \binom{n}{r} p_{s_0}^r p_{s_1}^{n-r} \mathcal{N}(\bar{e}_s; r\mu_0 + (n-r)\mu_1, r\sigma_0^2 + (n-r)\sigma_1^2).
\]

(4.22)

Herein and afterward, we show the PDF of random variable \( x \) having Gaussian distribution with mean and variance of \( \mu \) and \( \sigma^2 \), respectively, as \( \mathcal{N}(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \). When the number of active pulses \( n \) is not know, the PDF of the energy of received signal can be written as:

\[
f(\bar{e}_s|\mathcal{H}_1) = \frac{1}{1-(1-p_u)^m} \sum_{n=1}^{m} \binom{m}{n} p_u^n (1-p_u)^{m-n} f(\bar{e}_s|\mathcal{H}_1, n).
\]

(4.23)

For the case of binary modulation at the primary transmission, we can substitute (4.22) in (4.23) to get:

\[
f(\bar{e}_s|\mathcal{H}_1) = \frac{1}{1-(1-p_u)^m} \sum_{n=1}^{m} \sum_{r=0}^{n} \binom{m}{n} \binom{n}{r} p_u^n (1-p_u)^{m-n} p_{s_0}^r p_{s_1}^{n-r}
\times \mathcal{N}(\bar{e}_s; r\mu_0 + (n-r)\mu_1, r\sigma_0^2 + (n-r)\sigma_1^2).
\]

(4.24)

The probability of false alarm can be calculated based on the threshold \( \lambda \) on the received energy of \( m \) pulses as:

\[
P_F = \Pr(\bar{e}_s > \lambda|\mathcal{H}_0) = \sum_{i=0}^{M-1} p_{s_i} Q\left(\frac{\lambda - \bar{\mu}_i}{\bar{\sigma}_i}\right).
\]

(4.25)
\[
L(\bar{e}_s) = -\frac{1}{1-(1-p_u)^m} \sum_{n=1}^{m} \sum_{r=0}^{n} \binom{m}{n} \binom{n}{r} p_u^n (1-p_u)^{m-n} p_{s_0}^r p_{s_1}^{n-r} \frac{e^{-\frac{(\bar{e}_s-\bar{\mu}_0-(n-r)\mu_1)^2}{2\sigma_0^2+2(n-r)^2\sigma_1^2}}}{\sqrt{2\pi(n-r)^2\sigma_1^2}}.
\]

(4.28)

Also, the probability of detection can be written as:

\[
P_D = \Pr(\bar{e}_s > \lambda | \mathcal{H}_1) = \sum_{n=1}^{m} \sum_{r=0}^{n} \binom{m}{n} \binom{n}{r} p_u^n (1-p_u)^{m-n} p_{s_0}^r p_{s_1}^{n-r} \times \frac{1}{1-(1-p_u)^m} Q \left( \frac{\lambda - r\mu_0 - (n-r)\mu_1}{\sqrt{r^2\sigma_0^2 + (n-r)^2\sigma_1^2}} \right).
\]

(4.26)

Similar to single-pulse sensing, the optimal detection scheme can be applied to the received signal energy. Therefore, the test criterion for a given \( P_F \) that maximizes the probability of detection is:

\[
L(\bar{e}_s) = \frac{f(\bar{e}_s | \mathcal{H}_1)}{f(\bar{e}_s | \mathcal{H}_0)} \overset{\mathcal{H}_1}{\gtrsim} \gamma, \quad (4.27)
\]

where the likelihood ratio test \( L(\bar{e}_s) \) can be achieved by (4.28). Similar to single-pulse sensing, the optimal threshold can be chosen to minimize the probability of error as:

\[
\gamma = \frac{\Pr(\mathcal{H}_0)}{\Pr(\mathcal{H}_1)} = \frac{(1-p_u)^m}{1-(1-p_u)^m}. \quad (4.29)
\]

**Example:** The multiple-pulse sensing scheme is illustrated in Fig. 4.4(d), where the sensing nanodevice calculates the energy over four pulse durations, i.e., \( m = 4 \). The primary transmission (b) consists of two active symbols in \( S_1 \) and \( S_3 \) within the sensing window. The sensing nanodevice therefore considers this case as \( \mathcal{H}_1 \).
4.3 Asynchronous Molecular Channel Sensing

The single-pulse and multiple-pulses sensing detection schemes proposed in the previous section assume the perfect knowledge of primary timing at the sensing nanodevice. In this section, we consider the case that the timing information of the primary transmission is not available at the secondary. We assume that the energy detection begins in the middle of the symbol time with the delay of $\tau$ as shown in Fig. 4.4(e). For the sake analysis, we firstly calculate the PDF of the received energy where the sensing time $T_s$ is equal to the pulse width, i.e. $T_s = T_p$. Then, we generalize the analysis for smaller sensing times where $T_s < T_p$.

4.3.1 Pulse-Width Sensing Period

When $T_s = T_p$, the sensing starts and ends in two consequent symbol intervals, namely $S_1$ and $S_2$. In this case, hypothesis $H_0$ is considered when both the symbol durations are not in use by the primary. On the other hand, the hypothesis $H_1$ should be considered in three cases:

$$
H_1 : \begin{cases} 
C1 : & S_1 \text{ and } S_2 \text{ are used by primary,} \\
C2 : & \text{Only } S_1 \text{ is used by primary,} \\
C3 : & \text{Only } S_2 \text{ is used by primary.} 
\end{cases}
$$

(4.30)

For the case of $C1$, the energy of the received signal is the summation of the part of $S_1$ from $\tau$ to $T_p$ and the part of $S_2$ from 0 to $\tau$. Therefore, the received energy is the summation of two random variables both with mixture Gaussian distributions.
For calculating the PDF of the received energy in $C_1$, one only needs to convolve the PDF of two mixture Gaussian as follows:

$$f (e_s | C_1, \tau) = \sum_{i=0}^{M-1} \sum_{j=0}^{M-1} p_{s_i} p_{s_j} N (e_s; Q_i p (\tau, T_p) + Q_j p (0, \tau) \right)$$

$$+ Q_i p (\tau, T_p) \left[ 1 - p (\tau, T_p) \right] + Q_j p (0, \tau) \left[ 1 - p (0, \tau) \right] \right).$$

For the PDF of the received energy in $C_2$, we can write:

$$f (e_s | C_2, \tau) = \sum_{i=0}^{M-1} p_{s_i} N (e_s; Q_i p (\tau, T_p + \tau) \right)$$

$$+ Q_i p (\tau, T_p + \tau) \left[ 1 - p (\tau, T_p + \tau) \right] \right).$$

Likewise, for the case of $C_3$, assuming a free channel right before the sensing time (worse case scenario), the PDF of the energy can be written as:

$$f (e_s | C_3, \tau) = \sum_{i=0}^{M-1} p_{s_i} N (e_s; Q_i p (0, \tau), Q_i p (0, \tau) \left[ 1 - p (0, \tau) \right] \right).$$

For the case of $H_0$, when both $S_1$ and $S_2$ are not present, we can again consider the case where the previous symbol interval is used by the primary (worse case scenario). The PDF of the received signal energy can then be written as:

$$f (e_s | H_0, \tau) = \sum_{i=0}^{M-1} p_{s_i} N (e_s; Q_i p (\tau + T_p, 2T_p + \tau) \right)$$

$$+ Q_i p (\tau + T_p, 2T_p + \tau) \left[ 1 - p (\tau + T_p, 2T_p + \tau) \right] \right).$$

On the other hand, assuming that the probability of channel usage of channel by primary is $p_u$, the PDF of the received signal energy in the case of $H_1$ can be written
as:

\[ f(e_s|\mathcal{H}_1, \tau) = \] (4.35)

\[ p_u^2 f(e_s|C1, \tau) + p_u (1 - p_u) \left[ f(e_s|C2, \tau) + f(e_s|C3, \tau) \right]. \]

Considering the prior probability \( p_u \), we also have \( \Pr(\mathcal{H}_0) = (1 - p_u)^2 \) and \( \Pr(\mathcal{H}_1) = p_u^2 + 2p_u (1 - p_u) \). Using the Bayesian approach for unknown variable \( \tau \) [115], the ratio test becomes:

\[
L(e_s) = \frac{f(e_s|\mathcal{H}_1)}{f(e_s|\mathcal{H}_0)} = \frac{\int_{0}^{T_p} f(e_s|\mathcal{H}_1, \tau) f(\tau) \, d\tau}{\int_{0}^{T_p} f(e_s|\mathcal{H}_0, \tau) f(\tau) \, d\tau} \overset{\mathcal{H}_1}{\gtrless} \gamma. \] (4.36)

Assuming a uniform distribution for \( \tau \) over the interval \([0, T_p]\), the ratio test can be written as:

\[
L(e_s) = \frac{\int_{0}^{T_p} f(e_s|\mathcal{H}_1, \tau) \, d\tau}{\int_{0}^{T_p} f(e_s|\mathcal{H}_0, \tau) \, d\tau} \overset{\mathcal{H}_0}{\gtrless} \frac{(1 - p_u)^2}{p_u^2 + 2p_u (1 - p_u)} = \gamma. \] (4.37)

After some simple steps, the ratio test is simplified to:

\[ \frac{p_u^2 I_{C1} + p_u (1 - p_u) [I_{C2} + I_{C3}]}{I_{\mathcal{H}_0}} \overset{\mathcal{H}_0}{\gtrless} \gamma, \] (4.38)

where \( I_{C1}, I_{C2}, I_{C3}, I_{\mathcal{H}_0} \) are the integrals shown in (4.39), (4.40), (4.41) and (4.42).

4.3.2 Short Sensing Period

If we assume that the sensing period is less than the primary pulse width, i.e. \( T_s < T_p \), there are two possibilities for sensing. First, is the case that the sensing time spans over two primary transmission interval. This scenario is similar to \( \mathcal{H}_1 \) in Section IV.A, and we call it \( \mathcal{H}_1'' \), where the sensing time spans over two consecutive primary transmission intervals. On the other hand, since the sensing time is shorter than the
\[ I_{C1} = \sum_{i=0}^{M-1} \sum_{j=0}^{M-1} p_{si}p_{sj} \int_{0}^{T_p} \mathcal{N} (\epsilon; Q_p (\tau, T_p) \right. \\
+ Q_j p (0, \tau) , Q_i p (\tau, T_p) [1 - p (\tau, T_p)] + Q_j p (0, \tau) [1 - p (0, \tau)]) d\tau \\
I_{C2} = \sum_{i=0}^{M-1} p_{si} \int_{0}^{T_p} \mathcal{N} (\epsilon; Q_i p (\tau, T_p + \tau) , Q_i p (\tau, T_p + \tau) [1 - p (\tau, T_p + \tau)]) d\tau \\
I_{C3} = \sum_{i=0}^{M-1} p_{si} \int_{0}^{T_p} \mathcal{N} (\epsilon; Q_i p (0, \tau) , Q_i p (0, \tau) [1 - p (0, \tau)]) d\tau \\
I_{H_0} = \sum_{i=0}^{M-1} p_{si} \int_{0}^{T_p} \mathcal{N} (\epsilon; Q_i p (\tau + T_p, 2T_p + \tau) , Q_i p (\tau + T_p, 2T_p + \tau) \\
[1 - p (\tau + T_p, 2T_p + \tau)]) d\tau \]

pulse width and \( \tau \in [0, T_p] \), there might be a case where the whole sensing period resides in only one primary transmission interval. We call this case \( H_1' \). Therefore, \( H_1 \) is composed of two cases, i.e. \( H_1' \) and \( H_1'' \), and the distribution of energy given \( \tau \) in the case of \( H_1 \) can be written as:

\[ f (\epsilon|H_1, \tau) = f (\epsilon|H_1', \tau) Pr (H_1') + f (\epsilon|H_1'', \tau) Pr (H_1'') , \]

where

\[ Pr (H_1') = Pr (\tau < T_p - T_s) p_u = p_u (1 - \frac{T_s}{T_p}) , \]
and

\[ \Pr (H_1'') = \Pr (T_p - T_s \leq \tau \leq T_p) \left( p_u^2 + 2p_u(1 - p_u) \right) \]  (4.45)

\[ = (1 - \Pr (H_1')) \left( p_u^2 + 2p_u(1 - p_u) \right) \]

\[ = \frac{T_s}{T_p} \left( p_u^2 + 2p_u(1 - p_u) \right). \]

Furthermore, the PDF of the received energy in \( H_1' \) is:

\[ f (e_s|H_1', \tau) = \sum_{i=0}^{M-1} p_sN(e_s; Q_i p(\tau, T_s + \tau), Q_i p(\tau, T_s + \tau)[1 - p(\tau, T_s + \tau)]). \]  (4.46)

Similar to the \( H_1 \) case, we can divide \( H_0 \), into two cases, i.e. \( H_0' \) and \( H_0'' \), and then write the PDF of the received energy in this case as:

\[ f (e_s|H_0, \tau) = f (e_s|H_0', \tau) \Pr (H_0') + f (e_s|H_0'', \tau) \Pr (H_0''), \]  (4.47)

where \( \Pr (H_0') = (1 - \frac{T_s}{T_p})(1 - p_u) \) and \( \Pr (H_0'') = \frac{T_s}{T_p}(1 - p_u)^2 \). Considering the known prior probability of channel occupancy is available, the probability of hypothesis \( H_0 \) is then written as:

\[ \Pr (H_0) = \Pr (H_0') + \Pr (H_0'') = (1 - \frac{T_s}{T_p})(1 - p_u) + \frac{T_s}{T_p}(1 - p_u)^2, \]  (4.48)

and accordingly:

\[ \Pr (H_1) = \Pr (H_1') + \Pr (H_1'') = p_u(1 - \frac{T_s}{T_p}) + \frac{T_s}{T_p} \left( p_u^2 + 2p_u(1 - p_u) \right). \]  (4.49)

The PDF of the received energy for \( H_0' \) is calculated as:

\[ f (e_s|H_0', \tau) = \sum_{i=0}^{M-1} p_sN(e_s; Q_i p(\tau + T_p, T_p + T_s + \tau), Q_i p(\tau + T_p, T_p + T_s + \tau)[1 - p(\tau + T_p, T_p + T_s + \tau)]). \]  (4.50)
\[ I_{H_1'} = \sum_{i=0}^{M-1} p_s \int_0^{T_s} N(e_s; Q_i p(\tau, T_s + \tau), Q_i p(\tau, T_s + \tau) [1 - p(\tau, T_s + \tau)]) \, d\tau \]  

(4.54)

\[ I_{H_0'} = \sum_{i=0}^{M-1} p_s \int_0^{T_s} N(e_s; Q_i p(\tau + T_p, T_p + T_s + \tau), Q_i p(\tau + T_p, T_p + T_s + \tau) [1 - p(\tau + T_p, T_p + T_s + \tau)]) \, d\tau. \]

(4.55)

The final LRT detection test can be written based on the PDF of signals using Bayesian approach as:

\[
L(e_s) = \frac{\int_0^{T_s} f(e_s|H_1, \tau) \, d\tau \, H_1}{\int_0^{T_s} f(e_s|H_0, \tau) \, d\tau \, H_0} \gtrless \gamma, 
\]

(4.51)

where

\[
\gamma = \frac{(1 - \frac{T_s}{T_p})(1 - p_u) + \frac{T_s}{T_p}(1 - p_u)^2}{p_u(1 - \frac{T_s}{T_p}) + \frac{T_s}{T_p}(p_u^2 + 2p_u(1 - p_u))}. 
\]

(4.52)

The ratio test can then be simplified as:

\[
L(e_s) = \frac{p_u \left(1 - \frac{T_s}{T_p}\right) I_{H_1'} + I_{H_1''} \frac{T_s}{T_p}}{(1 - p_u) \left(1 - \frac{T_s}{T_p}\right) I_{H_0'} + I_{H_0''} (1 - p_u)^2 \frac{T_s}{T_p}} \gtrless \gamma, 
\]

(4.53)

where \(I_{H_1''}\) can be calculated from (4.35), (4.39), (4.40) and (4.41) when the upper limit of integral is changed to \(T_s\), and \(I_{H_1'}\) and \(I_{H_0'}\) are the integrals in (4.54) and (4.55).
4.4 Numerical Results

In this section, we evaluate the performance of the proposed sensing schemes for opportunistic cognitive molecular communication. All the results are achieved by the Monte Carlo simulations in MATLAB. For the performance of detection schemes, we employ the receiver operating characteristics (ROC) curves, i.e. $P_D$ versus $P_F$, and varying the detection thresholds. The area under an ROC curve denoted as AUC is considered a figure of merit in order to compare the sensing performance in different scenarios.

In the simulations, the diffusion coefficient is set to $D = 1 \text{ nm}^2/\text{ns}$ and the distance between the primary transmitter and receiver is $10 \mu\text{m}$. The pulse width of the primary network is $T_p = 0.8 \text{sec}$ and the distance between primary transmitter and the secondary sensing nanodevice is assumed to be $r_s = 30 \mu\text{m}$. We simulate the pulse-modulated signal of the primary nanonetwork with the maximum level of $Q_{M-1} = 10^4$ molecules, where $M$ is the size of modulation scheme at the primary nanonetwork. We assume that the number of released molecules for each symbol can be calculated by $Q_i = (i + 1)Q_{M-1}/M$ for $i = 0, \ldots, M - 2$. For the sake of simplicity of analysis, we also assume that the primary symbols are equally likely with the probability of $p_{s_i} = 1/M$. All the ROC curves except the last one are derived based on zero-velocity ($v = 0\text{m/s}$) of medium while the effect of this parameter is analyzed in Fig. 10.

The performance of the synchronous single-pulse sensing method is illustrated in Figs. 4.5 and 4.6 for different sensing window sizes, $\alpha$, and primary modulation
Figure 4.5: The ROC of the synchronous single-pulse sensing with simple thresholding of the received energy.

size, $M$. As discussed, we consider high probability of primary channel usage $p_u = 80\%$ in both figures. As seen in Fig. 4.5, the theoretical and experimental curves match perfectly. Fig. 4.5 also shows that the detection based on the energy of the received signal can provide a fixed probability of detection in some cases for a range of threshold values. On the other hand, the ROC curves in Fig. 4.6 show that the performance of the LRT detection is better than that of simple thresholding. The probability of detection is a strictly increasing function with respect to the threshold. As expected, the AUC is larger for the larger sensing window sizes as well as the smaller primary modulation sizes.

In Figs. 4.7 and 4.8, we show the performance of the proposed sensing schemes for the case of multiple-pulse sensing for different window sizes, $m$. It is assumed
that the primary uses a binary modulation and its probability of channel usage is $p_u = 20\%$. The first observation from these figures is that capturing the multiple primary pulses results in higher probability of detection for a fixed probability of false alarm. Fig. 4.7 shows that considering a larger window for capturing the energy for the ratio test does not necessarily provide a larger AUC while the curves in Fig. 4.8 show enhancement in the accuracy of sensing for larger window sizes. Therefore, the LRT based detection can significantly improve the performance of the sensing for synchronous multiple-pulse sensing scheme.

The performance of asynchronous sensing when the timing information of the primary is not available at the secondary nanonetwork is depicted in Fig. 4.9 for two different probability of channel usage $p_u = 20\%$ and $p_u = 80\%$. The comparison of
Figure 4.7: The ROC of the synchronous multiple-pulse sensing with simple thresholding of the received energy.

AUCs between the curves with different probability of usage shows the feasibility of the asynchronous sensing mechanism at low and high probability of channel usage. While the performance at low probability of usage is superior; lower probability of false alarm can be achieved for a fixed probability of detection. We can also observe that reducing the sensing window size does not deteriorate the detection performance significantly, while the size of the primary modulation notably affects it. This can be seen from Fig. 4.9 (see e.g. \( p_u = 0.8 \)) for a sufficiently small probability of false alarm. The probability of detection decreases by about 5% when the window size is changed from \( \alpha = 0.2 \) to \( \alpha = 0.8 \) for \( M = 2 \), while the difference between the cases of \( M = 2 \) and \( M = 8 \) is 25%.

The effect of velocity on the probability of error defined in (18) for syn-
Figure 4.8: The ROC of the synchronous multiple-pulse sensing with LRT detection method.

The synchronous single-pulse sensing LRT scheme is illustrated in Fig. 4.10. This figure is derived based on the optimal threshold value for $p_u = 0.8\%$ and $M = 4$. The figure shows that the probability of error is not a linear function of the velocity and the performance of the sensing is different for a range of medium velocity. For large values of drift velocity, the probability of error converges to zero. This result comes from the fact that the signal energy (number of received molecules) is mainly concentrated at the beginning of the sensing window for positive velocities, and the number of molecules arriving in $(T_p, (\alpha + 1)T_p)$ is generally smaller. This results in a reasonable difference between the probabilities of the two hypotheses, and thereby the error converges to zero.
Figure 4.9: The ROC of the asynchronous sensing for $p_u = 0.2$ and $p_u = 0.8$.

Figure 4.10: The probability of sensing error versus velocity; $M = 4$ and LRT based synchronous single-pulse sensing is considered.
CHAPTER V
SUMMARY AND FUTURE RESEARCH

This chapter provides a conclusion on the contributions of this dissertation. Also, we provide the areas of research which have potential for future work to follow on this work.

5.1 Conclusion

In this research work, different methods in employing cognitive radio capabilities were proposed in the presence of three emerging wireless technologies. In chapter one, we provided the background of cognitive radio network models and feasibility of scenarios considering three technologies.

In chapter two, we addressed the feasibility of opportunistic DoF usage of a $K$-user interference alignment network. We showed that for the proper number of antennas at cognitive radio transmitter and receiver, the secondary is able to utilize the unused DoFs of the primary system to transmit its own information while causing zero or minima interference to the primary receivers. We then came up with proper secondary precoding and decoding matrices to make the secondary transmission possible. We then proposed a two-stage spatial DoF sensing approach. Using our scheme, the CR receiver is able to quickly detect the availability of unused primary DoFs or
inactive streams. If a spatial hole is detected, in the second sensing stage, for finding the index set of inactive DoFs, the secondary system uses a spatial DoF index search method. With the help of simulations, we showed that the proposed opportunistic DoF usage scheme works well in practice and provides a significant throughput for the secondary system while causing no or minimal interference to the primary system.

In chapter three, we considered the system design and performance analysis of an overlay SM-based CR scheme that concurrently assists a primary system by relaying its information. While relaying the APM signal of the primary, the secondary transmitter is able to transmit its own information in the spatial domain without causing additional interference to the primary receiver. The primary receiver detects the APM symbols, while the secondary receiver estimates the index of the transmitting antenna to extract its own information. We analyzed the SER performance of such a scheme at both receivers and derived exact ASER expressions. We then determined asymptotically tight upperbounds for the ASER performance at high SNRs. Simulation results confirm our analytical derivations, and show that SM can be effectively used in overlay CR scenarios, providing an appealing transmission rate for the secondary network while assisting the primary transmission.

In chapter four, we proposed and analyzed a number of sensing schemes for cognitive molecular communication with the aim of cognitive radio concept. We derived novel sensing schemes for two different cases based on the availability of timing information of the primary nanonetwork at the secondary sensing nanodevice. Considering the concentration-based modulation scheme at the primary nanonetwork,
the proposed sensing schemes works based on the concept of energy detection by counting the number of received molecules at the secondary sensing nanodevice. In the case of synchronous sensing, we proposed two sensing schemes with short and long sensing windows and derived their probability of detection and false alarm. We also showed that the Bayesian approach can be applied to derive the likelihood ratio test when the cognitive nanonetwork is not synchronized with the primary nanonetwork. Numerical simulation of the proposed sensing schemes confirmed their effectiveness in sensing of molecular channels.

5.2 Future Research

While this dissertation has demonstrated the potential application of cognitive communications in three mostly recent technologies of wireless communications, many opportunities can be considered as the future work for extending the scope of this research and we summarize some of these suggestions as follows:

1. Although the proposed spatial sensing methods presented in Chapter 2 of dissertation have been designed for $K$-user interference channels, they can be also applied into other MIMO transmission schemes. Recently, massive MIMO wireless communications has been introduced which refers to the idea of equipping cellular base–stations with a very large number of antennas, and has been shown to potentially allow for orders of magnitude improvement in spectral and energy efficiency using relatively simple (linear) processing [116, 117]. The spatial sensing algorithm provided in this dissertation can be effectively considered for
massive MIMO application. We can consider a pair of MIMO transmitter and receiver close to each other as the secondary network are communicating with low power. If we assume that the precoder matrix and the channel knowledge of primary network is available at the secondary network, the secondary communication can be facilitated by the sensing mechanism proposed in Chapter 2. Therefore, the secondary receiver can detect the time slots that the primary base-station is not sending any data streams to a certain user close to the secondary receiver. Therefore, the secondary network can transmit opportunistically in the time slots when the closer user(s) is(are) not working while controlling the transmit power to be below the harmful threshold for other users.

2. The proposed cognitive spatial modulation scheme proposed in Chapter 3 has been studied for PSK modulation scheme. However, this scenario can be also tested on other modulation schemes such as QAM. Although it has been previously shown that PSK provides better error performance, the performances can be also analyzed for the proposed cognitive overlay scenario in which an additional amplified noise factor in the receivers affects the performance of the system.

3. The system model of cognitive spatial modulation scheme can be extended to more general scenarios. For instance, multiple secondary transmitters (or relays) can be considered. Also, the direct link between primary transmitter
and receiver can be applied to the system model. In this case, several scenarios can be studied with or without distributed beamforming in the second phase of relaying to enhance the performance of the primary while the secondary can send its own information.

4. If the idea of the proposed work in Chapter 4 about considering secondary networks in the molecular communications attracts the scientific community, this work can be considered as a beginning of a new research area in molecular communications. In a sense, several well-known techniques of cognitive radio in electromagnetic transmission can be applied in molecular communications. For example, the feasibility of considering other scenarios such as overlay and underlay approaches can be effectively studied. Also, we considered a basic sensing (detection) method for the cognitive molecular communications in Chapter 3 to better analyze the feasibility of the proposed scenario. Therefore, other advanced sensing mechanisms can be used for this purpose.

5. In Chapter 4, we studied the cognitive capability in molecular communications for a concentration-based modulation scheme. However, this scenario can be properly used for other modulation schemes such as time-based modulations. Therefore, we recommend the study of cognitive molecular communications on other recently published transmission schemes to better show the application of opportunistic usage of the molecular channels in other practical scenarios.
APPENDICES
APPENDIX A

PROOF OF LEMMA 3

To see the effect of the number of secondary receive antennas, $N^{[0]}$, on the average SINR, we show that by adding one additional antenna at the secondary receiver, the average SINR increases. We first define the new vector $\tilde{H}_l$ and matrix $\tilde{B}_l$ of the receiver with one additional antenna as:

$$ \tilde{H}_l = \begin{bmatrix} H_l \\ h_l \end{bmatrix}, \quad \tilde{B}_l = \begin{bmatrix} B_l & B_1 \\ B_1^\dagger & b_2 \end{bmatrix}, \quad (A.1) $$

where $\tilde{H}_l$ and $H_l$ are vectors of size $(N^{[0]} + 1) \times 1$ and $N^{[0]} \times 1$, respectively, and matrices $\tilde{B}_l$ and $B_l$ are of size $(N^{[0]} + 1) \times (N^{[0]} + 1)$ and $N^{[0]} \times N^{[0]}$, respectively. Hence, the inverse of partitioned matrix $\tilde{B}_l$ can be written as:

$$ \tilde{B}_l^{-1} = \begin{bmatrix} \hat{B}^{-1} & \hat{B}_1 \\ \hat{B}_1^\dagger & \hat{b}_2 \end{bmatrix}, \quad (A.2) $$

where $\hat{B} = B_l - B_1 B_1^\dagger$, and $\hat{B}_1$ and $\hat{b}_2$ are well-defined vector and scalar, respectively [78]. The average SINR at the secondary receiver in (2.23) can be written as:

$$ \mathbb{E} [\gamma^{\text{CR max}}_l |_{(N^{[0]} + 1)}] = \frac{p^{[0]}}{d^{[0]}} \mathbb{E} [\tilde{H}_l^\dagger \tilde{B}_l^{-1} \tilde{H}_l] \quad (A.3) $$

$$ = \frac{p^{[0]}}{d^{[0]}} \left( \mathbb{E} [H_l^\dagger \hat{B}^{-1} H_l] + 2 \Re \{ \mathbb{E} [h_l^* \hat{B}_1^\dagger H_l] \} + \mathbb{E} [|h_l|^2 \hat{b}_2] \right). $$

Considering single stream transmission by the secondary, and based on the definition of interference matrix $B_l$ in (10), there is no common random variable in $B_l$ and
Therefore, $h_l$, $h_l$, and $\hat{B}_1$ are independent, and the second term in (A.3) is zero because $\mathbb{E}[h_l] = 0$. Furthermore, the third term in (A.3) is always positive. To prove the lemma, it suffices to compare the first term in this equation with the average SINR in the case of $N^{[0]}$ number of antennas.

For this purpose, we first prove that $\hat{\mathbf{B}}$ is positive definite (PD); i.e. for all non-zero vectors $X$, $X^\dagger \hat{\mathbf{B}} X > 0$. Considering that matrix $\tilde{\mathbf{B}}_l$ is PD, for any given non-zero vector $X^H = \begin{bmatrix} X_1^H & x_2^* \end{bmatrix}$:

$$X^\dagger \tilde{\mathbf{B}}_l X = X_1^\dagger \left( b_2 \mathbf{B}_l - B_1 B_1^\dagger \right) X_1 + \|B_1^\dagger X_1 + b_2 x_2\|^2.$$  \hspace{1cm} (A.4)

By choosing $x_2 = -B_1^\dagger X_1/b_2$, it follows that the first term in (A.4), i.e.

$$X_1^\dagger \left( b_2 \mathbf{B}_l - B_1 B_1^\dagger \right) X_1$$  \hspace{1cm} (A.5)

is always positive. Therefore, the matrix $\left( b_2 \mathbf{B}_l - B_1 B_1^\dagger \right)$ is PD and hence, $\hat{\mathbf{B}}$ is PD. Considering the fact that $\mathbf{B}_l \succeq \hat{\mathbf{B}}$ and since both matrices are PD, we can conclude that $\mathbf{B}_l^{-1} \succeq \hat{\mathbf{B}}^{-1}$ \cite{32}. The first term in (A.3) is always larger than the average SINR when the secondary receiver has $N^{[0]}$ antennas. So, $\mathbb{E}[\tilde{\gamma}_l^{\text{CRmax}}]_{(N^{[0]}+1)} \geq \mathbb{E}[\gamma_l^{\text{CRmax}}]_{(N^{[0]})}$, and the average SINR is an increasing function of the number of antennas at the secondary receiver.
APPENDIX B

PROOF OF LEMMA 5

The first constraint of (2.49) implies that the sensing vector $D_i^{[i]}$ must be orthogonal onto the subspace of matrix $R_i^{[i]}$ in $N^{[0]}$-dimensional space of the secondary receiver. The orthogonal projection of the desired vector $H^{[0i]}V_{[l]}^{[i]}$ onto subspace $R_i^{[i]}$ is [88, 5.13]:

$$C_i^{[i]} = R_i^{[i]} \left( R_i^{[i]\dagger} R_i^{[i]} \right)^{-1} R_i^{[i]\dagger} H^{[0i]} V_{[l]}^{[i]}.$$  \hspace{1cm} (B.1)

On the other hand, the objective function of (2.49) is to find an orthogonal vector to $C_i^{[i]}$ with maximum inner product with the desired vector $H^{[0i]} V_{[l]}^{[i]}$. Therefore, the normalized vector $D_i^{[i]}$ should be in the direction of the orthogonal projector onto subspace $R_i^{[i]}$ and can be written as:

$$D_i^{[i]} = \frac{1}{\lambda} \left( H^{[0i]} V_{[l]}^{[i]} - C_i^{[i]} \right),$$  \hspace{1cm} (B.2)

where normalization parameter $\lambda$ is:

$$\lambda^2 = \left( V_{[l]}^{[i]} H^{[0i]\dagger} - C_i^{[i]\dagger} \right) \left( H^{[0i]} V_{[l]}^{[i]} - C_i^{[i]} \right).$$  \hspace{1cm} (B.3)
APPENDIX C

PROOF OF THEOREM 1

The PDF of the ratio of these two random variables can be written as [118, 6.43]:

\[ f_Z(z) = \int_0^\infty (z_2 + C) f_{Z_1,Z_2}((z_2 + C)z, z_2) \, dz_2, \quad (C.55) \]

By substituting (3.23), and after some manipulations, the PDF is derived as:

\[ f_Z(z) = e^{-Cz(1-\rho)} \frac{\mu_1}{\mu_1 \mu_2 (1-\rho)} \int_0^\infty z_2 \exp \left[-\frac{(\mu_1 + \mu_2 z)}{\mu_1 \mu_2 (1-\rho)} z_2 \right] I_0 \left[ \frac{2 \sqrt{\rho z_2}}{(1-\rho) \mu_1 \mu_2 \sqrt{z_2 (z_2 + C)}} \right] \, dz_2 \]

\[ + \frac{Ce^{-Cz(1-\rho)} \mu_1}{\mu_1 \mu_2 (1-\rho)} \int_0^\infty \exp \left[-\frac{(\mu_1 + \mu_2 z)}{\mu_1 \mu_2 (1-\rho)} z_2 \right] I_0 \left[ \frac{2 \sqrt{\rho z_2}}{(1-\rho) \mu_1 \mu_2 \sqrt{z_2 (z_2 + C)}} \right] \, dz_2 \]

(C.56)

The second term in (C.56) can be written as the Laplace transform of the Modified Bessel function of the first kind, and can be calculated as:

\[ \int_0^\infty e^{-sz_2} I_0 \left[ b(z) \sqrt{z_2 (z_2 + C)} \right] \, dz_2 = \mathcal{L} \left\{ I_0 \left[ b(z) \sqrt{t(t + C)} \right] \right\} \]

\[ = \frac{1}{\sqrt{s^2 - b(z)^2}} \exp \left[ \frac{C}{2} \left( s - \sqrt{s^2 - b(z)^2} \right) \right], \quad (C.57) \]

where \( b(z) = \frac{2 \sqrt{\rho z}}{(1-\rho) \sqrt{\mu_1 \mu_2}} \). For calculating the first term in (C.56), we rely on Laplace transform. The result can be obtained by taking the derivative of \( \mathcal{L} \left\{ I_0 \left[ b(z) \sqrt{t(t + C)} \right] \right\} \)

which can be written as:

\[ \int_0^\infty z_2 e^{-sz_2} I_0 \left[ b \sqrt{z_2 (z_2 + C)} \right] \, dz_2 = e^{\frac{C}{2} (s-\beta(z))} \left[ -C \frac{1}{2 \beta(z)^2} (\beta(z) - s) + \frac{s}{\beta(z)^3} \right], \quad (C.58) \]

where \( s = \frac{(\mu_1 + \mu_2 z)}{\mu_1 \mu_2 (1-\rho)} \) and \( \beta(z) = \sqrt{s^2 - b^2} \). This finalizes the proof.
By the definition of correlation coefficient, we have:

\[
\rho_{X_1,X_2} = \frac{\mathbf{E}(X_1X_2) - \mathbf{E}(X_1)\mathbf{E}(X_2)}{\sqrt{\mathbf{E}(X_1^2) - \mathbf{E}(X_1)^2}\sqrt{\mathbf{E}(X_2^2) - \mathbf{E}(X_2)^2}},
\]  

(D.1)

By substituting \(X_1\) and \(X_2\), and considering the fourth moment of \(Y\) as \(\mathbf{E}(|Y|^4) = 2\sigma_Y^4\), the first and second terms in the numerator can be, respectively, written as:

\[
\mathbf{E}\left(|Y^{(p)} - W^{(p')}|^{2}\left(|Y|^2 + C\right)\right)
\]  

(D.2)

\[
= \left|x^{(p)}\right|^2 2\sigma_Y^4 + \left|x^{(p')}\right|^2 \sigma_{W}\sigma_{Y}^2 + \left|x^{(p)}\right|^2 C\sigma_Y^2 + \left|x^{(p')}\right|^2 \sigma_{W}^2 C,
\]

By subtracting these two terms, the numerator of (D.1) is \(\left|x^{(p)}\right|^2 \sigma_Y^4\). Similarly, the first and second terms in the denominator in the right side of (D.1) can be, respectively, written as:

\[
\sqrt{\mathbf{E}\left(|Y^{(p)} - W^{(p')}|^4\right) - \mathbf{E}\left(|Y^{(p)} - W^{(p')}|^2\right)^2}
\]  

(D.3)

\[
= \sqrt{\mathbf{E}\left(|Y^{(p)} - W^{(p')}|^4\right) - \left(|x^{(p)}|^2\sigma_Y^2 + |x^{(p')}|^2\sigma_W^2\right)^2}
\]

\[
= \left|x^{(p)}\right|^2 \sigma_Y^2 + \left|x^{(p')}\right|^2 \sigma_W^2,
\]

and

\[
\sqrt{\mathbf{E}\left((|Y|^2 + C)^2\right) - \mathbf{E}(|Y|^2 + C)^2} = \sigma_Y^2.
\]  

(D.4)

So, the denominator in (D.1) is \(\left|x^{(p)}\right|^2 \sigma_Y^4 + \left|x^{(p')}\right|^2 \sigma_W^2 \sigma_Y^2\), and this Proves the lemma.
APPENDIX E

PROOF OF LEMMA 8

Considering that the values of the Associated Legendre functions in (3.36) are the same, for brevity, we replace them with a new variable $\eta_i$, which is defined as:

$$
\eta_i = \frac{a_i}{2} \left( \lambda r^2 + \frac{a_i^2}{4} \right)^{-\frac{1}{2}}.
$$

(E.1)

By substituting $\eta_i$ from (E.1) into the exact average PEP for $i = 1, 3, 4$ in (3.36), we have:

$$
P_{EP_i} = \frac{a_i\sigma_{SR}^2}{4a_i} + \lambda - \frac{5\lambda \lambda' \lambda^2}{4a_i^3} \binom{3}{2} F_2 \left(1, \frac{7}{4}, \frac{9}{4}, \frac{3}{2}, \frac{5}{2}; -\frac{4\lambda'^2}{a_i^2} \right)
$$

$$
- \frac{3\lambda}{8} \frac{1}{a_i} \eta_i \frac{3}{2} Q_{1/2}^{-1} [\eta_i] + \lambda \lambda' \frac{1}{4a_i^2} \binom{3}{2} F_2 \left(1, \frac{5}{4}, \frac{7}{4}, \frac{3}{2}, \frac{3}{2}; -\frac{4\lambda'^2}{a_i^2} \right)
$$

$$
+ \frac{\lambda}{4} \frac{1}{a_i} \eta_i \frac{3}{2} Q_{0}^{-1} [\eta_i].
$$

(E.2)

Our aim is to find an approximation for $P_{EP_i}, i = 1, 3, 4$, at high SNRs (i.e. large $a_i$ values). Since for the Generalized Hypergeometric Series $\binom{3}{2} F_2 (\cdot, \cdot; \cdot; \cdot; x) = 1$ when $x \to 0$, at high SNRs, we have:

$$
\binom{3}{2} F_2 \left(1, \frac{7}{4}, \frac{9}{4}, \frac{3}{2}, \frac{5}{2}; -\frac{4\lambda'^2}{a_i^2} \right) = \binom{3}{2} F_2 \left(1, \frac{5}{4}, \frac{7}{4}, \frac{3}{2}, \frac{3}{2}; -\frac{4\lambda'^2}{a_i^2} \right) = 1.
$$

(E.3)

Obviously, $\lim_{a_i \to \infty} \eta_i = 1^-$, and we can consider the behavior of the Associated Legendre function [119, 14.8.6] to approximate $Q_{1/2}^{-1} [\eta_i]$ at high SNRs as:

$$
Q_{1/2}^{-1} [\eta_i] \approx \frac{2}{3} \left( \frac{2}{1 - \eta_i} \right)^{1/2}.
$$

(E.4)
Also, considering the approximation of the Associated Legendre functions of zero order [119, 14.8.3], the second Associated Legendre function in (E.2) can be written as:

\[ Q^0 \frac{1}{2} [\eta_i] = \frac{1}{2} \ln \left( \frac{2}{1 - \eta_i} \right) - \gamma - \psi \left( \frac{3}{2} \right) + \mathcal{O} (1 - \eta) \quad \text{(E.5)} \]

\[ \approx \frac{5}{2} \ln (2) - 2 - \frac{1}{2} \ln (1 - \eta_i) , \]

where \( \gamma \) is the Euler-Mascheroni constant, and \( \psi(\cdot) \) is the Digamma function. Substituting (E.3), (E.4) and (E.5) in (E.2) provides the approximation of \( \overline{\text{PEP}}_i \) at high SNRs as:

\[
\overline{\text{PEP}}_i = \frac{a_i \sigma^2_{SR} + \lambda}{4a_i} - \frac{5\lambda \lambda'}{4a_i} + \frac{3\lambda \lambda'}{4a_i^2} - \frac{\lambda}{4a_i} \eta_i^2 \left( \frac{2}{1 - \eta_i} \right)^{\frac{1}{2}} \\
+ \frac{\lambda}{4a_i} \left( \frac{5}{2} \ln (2) - 2 - \frac{1}{2} \ln (1 - \eta_i) \right) , \quad i = 1, 3, 4.
\quad \text{(E.6)}
\]

We can further simplify (E.6) by considering:

\[
\lim_{a_i \to \infty} \frac{\lambda}{4a_i} \left( \frac{5}{2} \ln (2) - 2 - \frac{1}{2} \ln (1 - \eta_i) \right) = \frac{1}{4} \lambda \lambda' . \quad \text{(E.7)}
\]

Considering \( \lim_{a_i \to \infty} (1 - \eta_i) = \frac{2\lambda^2}{a_i^2} \) and with some simple manipulations, the lemma is proved.


