Spectrum Sensing and Interference Mitigation in Cognitive Radio Networks

Von der Fakultät für Elektrotechnik und Informationstechnik
der Rheinisch–Westfälischen Technischen Hochschule Aachen
zur Erlangung des akademischen Grades
einer Doktorin der Ingenieurwissenschaften
genehmigte Dissertation

vorgelegt von
Xitao Gong, M.Sc.
aus Nanjing, China

Berichter: Universitätsprofessor Dr.-Ing. Gerd Ascheid
Universitätsprofessor Dr. Petri Mähönen

Tag der mündlichen Prüfung: 21.02.2014

Abstract

One concept to increase spectral efficiency is dynamic spectrum access (DSA). The secondary users coexist with the primary users in the same radio-frequency spectrum. Such shared usage demands for an interference avoidance or interference mitigation at the secondary users. This is especially challenging due to the limited cooperation between the primary users (PUs) and the secondary users (SUs). Motivated by this fact, the focus of this thesis is a comprehensive study on interference management strategies and the characterization of the achievable performance of secondary systems.

Spectrum sensing aims at detecting the presence or absence of the PUs. The main challenge encountered is the high requirement on sensitivity, reliability, and agility, especially in case of incomplete knowledge of the transmission channels. Therefore, the SUs need to efficiently utilize the limited a priori knowledge related to the primary transmission to improve the sensing performance. In this thesis, the generalized likelihood ratio test framework is applied to cooperative sensing problems with an unknown structure of the primary signal space and unknown noise variances at the SUs. The efficiency of the resulting spectrum sensing algorithms is demonstrated as well as the effectiveness in countering the "hidden primary user" problem.

Based on limited knowledge related to the primary transmission, the SUs' transceiver strategies are optimized in order to achieve the tradeoff between improved secondary network throughput and, most critically, constrain the performance loss of the primary transmission. For a single-antenna spectrum sharing system, the power allocation strategies are investigated for the SUs subject to different quality of service constraints on the primary link. Not only optimal and low-complexity near-optimal power allocation strategies are developed, but also the achievable performance of the system is approximately evaluated in closed form. Additionally, for multi-antenna spectrum sharing networks, efficient transceiver optimization strategies are developed under the consideration of imperfect channel state information. The robustness, optimality, and convergence behavior of the different proposed algorithms are quantitatively verified and compared.

The essential "cognitive" property of DSA in cognitive radio networks consists of two aspects: the acquirement of the useful information from the environment and the

utilization of such information to improve the spectrum efficiency. This is demonstrated with the study of a hybrid paradigm, in which the SU exploits the spectrum sensing and location information to adapt the transmit power level. Compared to the standard paradigms, e.g., opportunistic transmission and spectrum sharing without sensing, the proposed strategies in the hybrid paradigm achieve better performance.

Kurzfassung

Ein Konzept zur Erhöhung der spektralen Effizienz ist der dynamische Zugang zum Spektrum (Dynamic Spectrum Access, DSA). Sekundäre Nutzer koexistieren dabei mit den primären Nutzern eines Frequenzbands. Eine solche gemeinsame Nutzung setzt Interferenzvermeidung bzw. Interferenzminimierung durch die sekundären Benutzer voraus. Dies ist insbesondere durch die begrenzte Kooperation zwischen den primären Benutzern (Primary Users, PU) und den sekundären Nutzern (Secondary Users, SU) eine Herausforderung. Schwerpunkt dieser Arbeit ist eine umfassende Untersuchung möglicher Interferenz-Management-Strategien und die Charakterisierung der erzielbaren Performance von sekundären Systemen.

Durch Spektrum-Sensing kann die Anwesenheit oder Abwesenheit der PUs detektiert werden. Die größten Herausforderungen hierbei sind die hohen Anforderungen an Empfindlichkeit, Zuverlässigkeit und Flexibilität, vor allem bei unvollständiger Kenntnis der Übertragungskanäle. Daher müssen die SUs die begrenzte a priori Kenntnis bzgl. der primären Übertragung effizient nutzen, um eine ausreichende Sensing-Performance zu erreichen. In der Dissertation wird hierfür das verallgemeinerte Likelihood-Ratio-Test-Framework auf kooperative Sensing-Probleme mit einer unbekannten Struktur des primären Signalraums und unbekannter Rauschvarianzen an den SUs angewandt. Die Effizienz der resultierenden Spektrum-Sensing Algorithmen für Fading-Kanäle wird auch in Bezug auf das „hidden primary user"-Problem gezeigt.

Basierend auf einer begrenzten Kenntnis der primären Übertragung, werden die Transceiver-Strategien der SUs optimiert, um einen Kompromiss zwischen hohem sekundären Netzwerk-Durchsatz und der Beschränkung des Performance-Verlusts der primären Übertragung zu erreichen. Für Spektrum-Sharing-Systeme mit Einzel-Antennen-Geräten werden die Strategien zur Leistungsallokation an den SUs unter der Bedingung verschiedener „Quality of Service" Einschränkungen für die primäre Verbindung untersucht. Es werden nicht nur optimale Strategien und nahezu optimale Strategien geringer Komplexität entwickelt, sondern es wird auch die erzielbare Performance des Systems näherungsweise in geschlossener Form angegeben. Zusätzlich werden für Mehrantennen-Spektrum-Sharing-Systeme effiziente Transceiver-Strategien unter Berücksichtigung der unvollkommenen Kanalkenntnis entwickelt. Die

Robustheit, die Optimalität und das Konvergenzverhalten der verschiedenen vorgeschlagenen Algorithmen werden quantitativ bestätigt und verglichen.

Die essentielle „kognitive“ Eigenschaft von DSA in Cognitive Radio Networks besteht aus zwei Aspekten: Das Erfassen nützlicher Informationen aus der Umwelt und die Nutzung dieser Informationen, um die spektrale Effizienz zu erhöhen. Dies wird beispielhaft mit einer Studie eines Hybrid-Paradigmas demonstriert, bei dem die SUs die Sendeleistung in Abhängigkeit der Spektrum-Sensing-Informationen und von Orts-Informationen anpassen. Im Vergleich zu den Standard-Paradigmen, d.h. der opportunistischen Übertragung oder der gemeinsamen Nutzung des Frequenzspektrums ohne Sensing, erreichen die vorgeschlagenen Strategien im Hybrid-Paradigma eine bessere Performance.

Contents

Chapter 1

Introduction

Wireless networks are currently assigned by spectrum regulatory bodies to license holders for exclusive usage. Due to the ever-increasing demand for wireless systems and services, such static spectrum allocation policy yields a severely limed available wireless spectrum. In contrast, recent observations and measurement studies of some regulatory agencies such as the Federal Communications Commission (FCC) have shown that most of the radio frequency spectrum was inefficiently utilized [1]. The paradox between spectrum scarcity and spectrum under-utilization motivates the initial idea of cognitive radios (CRs) which enable dynamic spectrum access (DSA) in licensed bands [90].

A CR is an intelligent device that is aware of the environment, adapts its transceiver parameters accordingly, and transmits in licensed bands with limited interference to licensed users. Due to the lack of priority to access the licensed spectrum resources, unique challenges are imposed to enable DSA, such as interference avoidance with licensed networks, quality of service (QoS) awareness for dynamic and heterogeneous transmission, and seamless communication irrelevant to the variations of the licensed users' activity [5, 6]. To address these challenges, two main characteristics of CRs are introduced in [55]

- *Cognitive capability*: A CR system is able to obtain the information from its radio environment and keep on tracking its variations. Such information is used to characterize favorable conditions in the environment and choose appropriate operating parameters for the CR devices.

- *Reconfigurability*: A CR system is able to be dynamically programmed concerning its software or hardware settings based on the transmission decisions made according to distinct scenarios.

The CR users are usually referred to as *the secondary users* (SUs), while the incumbent users in the licensed spectrum band are termed *the primary users* (PUs).

New functionalities in a spectrum management framework are required for CR networks, such as spectrum sensing, spectrum decision, spectrum sharing, and spectrum mobility [5]. These functions are realized through a cross-layer design approach. Since the focus of this research is on the PHY layer design of CR systems, two critical functions are considered. First, *spectrum sensing* which indicates that a CR system needs to monitor the available spectrum in time and space. Based on the observations, the spectrum opportunities are identified and their characteristics are captured. Second, *spectrum sharing* which means that a CR system is required to coordinate the

transmission and the access to licensed bands, and to avoid severe interference to other users.[1]

Although the initial concept of CR systems is to enable the DSA in licensed bands [2], this idea can also be applied to the unlicensed bands, e.g., in the 2.4 GHz unlicensed band which possibly hosts systems like Bluetooth, Wi-Fi, etc [38]. The "intelligence" of the SUs enables them to be aware of the spectrum occupancy and adapt their transmission in accordance with informative knowledge learned from the environment. Hence, less interference and improved spectral efficiency are envisioned.

The standardization of CR systems has been recently carried on by several organizations, including the International Telecommunication Union (ITU), the Institute of Electrical and Electronics Engineers (IEEE), the European Telecommunications Standards Institute (ETSI), etc. For example, the IEEE 802.22 Wireless Regional Area Network (WRAN) standard allows the usage of CR techniques in the very high frequency and ultra high frequency (VHF/UHF) bands licensed to television (TV) broadcasting services and other coexisting services such as wireless microphones. The work [31] provides an overview of recently developed international standards related to CR systems.

1.1 Research Objectives

DSA is a promising technology to greatly improve the spectrum efficiency in wireless communication systems. However, new challenges encompass the design of radio devices with "full cognition". The essential requirement is to properly respond to the interaction with the operating environment and avoid the harmful interference to both licensed and unlicensed users. Hence, the ultimate objective of this research is to investigate several key interference management techniques in the DSA environment. More specifically, we focus on addressing the following topics.

1.1.1 Spectrum Sensing

Spectrum sensing is a basic functionality for the realization of the CR technology through interference avoidance. Driven by the initial motivation of CR systems to opportunistically use the available licensed bands, spectrum sensing identifies the spectrum access opportunities for the SUs to avoid interfering with the PUs' transmission. Such opportunities can be characterized by the PUs' activity information. The main challenge of spectrum sensing is a rigorous requirement on agility, sensitivity, and reliability, i.e, the SUs need to detect the PUs' activity statuses quickly in a low signal to noise ratio (SNR) region but with high accuracy. For example, the IEEE 802.22 WRAN standard requires the detection of the incumbent TV broadcasting signals for power as low as -116 dBm and wireless microphone signals for power

[1] Here, the meaning of spectrum sharing is slightly different from the definition in [5] since it is not restricted to spectrum sharing only between the CR users but also allows for the coexistence with primary systems.

as low as -107 dBm. The sensing procedure should be finished in less than 2 seconds with at least 90% probability of detection and at most 10% false alarm rate [3, 118]. Such requirements become even more challenging when facing the hidden primary user problem: for example, if a SU is located in a deep fade of the primary transmission, it might cause unwanted interference to the PU due to a missed detection of the PUs' transmission [139]. One effective method, termed cooperative spectrum sensing that combines the sensing observations from multiple SUs, ameliorates the sensing performance by exploiting more degrees of freedom (DoFs) of the SUs.

The environment uncertainties at the SUs, including the fading effect of the channels and the noise uncertainty [113, 124], may severely degrade the sensing performance, even in a cooperative manner. Furthermore, a complete a priori knowledge for detection, e.g., the channel state information (CSI) and the information concerning the primary signals, is hard to be perfectly obtained at the SUs. In order to address these problems, the focus of the present investigation lies in the effective exploitation of the available but limited information at the SUs to optimize the cooperative spectrum sensing approaches.

1.1.2 Inteference Mitigation

Apart from opportunistic transmission, spectrum sharing, which allows for the concurrent transmission of the PUs and the SUs, also has the potential to achieve high spectral efficiency as long as the interference caused to the primary system and among the SUs is properly controlled [38, 146]. Moreover, spectrum sharing might occur in opportunistic transmission under a missed detection of the primary transmission.

The network model for spectrum sharing can be considered as interference channels consisting of the users with a hierarchical priority. Thus, different from traditional multiuser networks, the new challenge involved in DSA is the interference mitigation between the primary and secondary systems. Due to limited cooperation between the PUs and the SUs, the SUs only have restricted knowledge of the primary system which makes interference management more difficult. To deal with this problem, the CR system can perform dynamic resource allocation, such as the proper selection of the frequency bands, adjustment of the transmit power, or transceiver optimization, to achieve QoS requirements and constrain the interference.

As a basis to optimize the dynamic resource allocation, we need to choose appropriate performance measures to evaluate the performance degradation to primary systems. The commonly used measure is the so called "interference temperature" [33] representing the received interference power at the primary receivers (PRs). Alternatively, limiting the primary rate loss is the ultimate goal in regulating the secondary transmission. Thus, the performance loss of the PUs can also be assessed with rate-related measures, e.g., an outage probability of the primary link. According to the type of the available CSI at the SUs, the aforementioned metrics can be constrained either in a long-term or a short-term manner.

Power control is a simple but very effective way for interference mitigation. We are interested in assessing the achievable performance regarding the secondary transmis-

sion given different QoS constraints on primary systems and limited a priori knowledge at the SUs. From a practical point of view, low-complexity and efficient near-optimal strategies are also desirable. In addition, in realistic systems, the PRs are sometimes passive receivers or mobile receivers, e.g., in TV broadcasting systems or in mobile cellular networks, respectively. Hence, their location information is difficult to be obtained at the SUs. Such location uncertainty is also required to be taken into consideration in the interference modeling. Furthermore, sensing results can also be incorporated in the design of the power allocation.

Nowadays, the use of multiple antennas at the transceivers is well established due to the enhancement of the spectral efficiency by exploiting the spatial domains. By steering the directional transmission at the targeted users and avoiding it towards the interfering users, transceiver optimization achieves the tradeoff between interference mitigation and improving the performance of the desired links. In reality, perfect CSI of the entire network is hard to obtain at the SUs. Considering this effect in the algorithmic design usually results in intractable non-convex optimization problems. Hence, one focus of this thesis is to develop efficient robust transceiver optimization algorithms in cognitive networks.

1.2 Outline

The target of this thesis is to investigate interference management techniques in CR networks with focuses on cooperative spectrum sensing strategies and interference mitigation strategies for spectrum sharing. On the one hand, regarding cooperative sensing, we study the effective utilization of the limited a priori knowledge regarding the primary transmission and noise variances at the SUs. On the other hand, concerning interference mitigation in spectrum sharing systems, we consider exploiting dynamic resource allocation while taking into account the realistic challenge that the SUs only have partial CSI related to the primary transmission. Specifically, in single-antenna spectrum sharing systems, the relation between the achievable performance of the secondary transmission and different QoS constraints of the primary transmission is addressed. To further exploit the spatial DoFs introduced by using multiple-antenna transmission, we proceed to develop efficient transceiver optimization strategies for interference mitigation. Finally, the question on how to explore the learning functionality in the design of transmission strategies at the SUs is answered by an exemplified investigation on sensing-based power allocation algorithms.

The structure of the thesis is as follows. In Chapter 2, three paradigms of CR systems are introduced: the *interweave*, the *underlay*, and the *overlay* paradigm. Briefly speaking, the interweave paradigm aims at guaranteeing the exclusive transmission of the PUs and the SUs, the underlay paradigm mandates the simultaneous transmission of both kinds of users, whereas the overlay paradigm also permits the concurrent transmission of both kinds of users by allowing the SUs to use part of the secondary transmission resources to relay the primary message. The preliminary assumptions and the a priori information required for each paradigm are addressed, followed by a

brief discussion on the realistic solutions to deal with the respective challenges. Based on the discussion, spectrum sensing and dynamic resource allocation are shown to be two key techniques for interference management in CR systems. Therefore, we recapitulate commonly-used spectrum sensing strategies with a focus on the approaches based on limited a priori knowledge. Additionally, ideas and abstract models for dynamic resource allocation in CR systems are addressed including a detailed literature review.

Cooperative spectrum sensing algorithms with limited a priori information regarding the primary transmission and noise variances are elaborated in Chapter 3. We are interested in how to exploit the available but limited information to improve the sensing performance. Since such problems are modeled as binary hypothesis testing problems with unknown parameters, the generalized likelihood ratio test (GLRT) framework is applied to derive the solutions. Firstly, assuming an unknown structure of the primary signal space at the SUs, we exploit the rank information of the primary signal space, which represents the DoFs of signals, to estimate its structure and design the GLRT-based methods. Secondly, we consider the influence of the fading sensing and reporting channels on the cooperative sensing. The underlying assumption is that the SUs only have partial CSI of the sensing channels and do not know the structure of the primary signal space. The GLRT-based algorithms are proposed under such circumstances to ameliorate the performance degradation caused by environment uncertainties, e.g., the fading effects and noise variance uncertainty.

In Chapter 4, power allocation strategies for the secondary transmission are investigated subject to different QoS constraints on the primary link. The goal is to characterize the connection between the achievable gain of the SUs and the tolerable performance loss of the PUs. Assuming only partial CSI related to the PR is available at the SU, we study the power allocation problems constrained by two kinds of constraints on the primary transmission: a conventional interference temperature (IT) constraint and an outage probability constraint. We not only address the optimal power allocation strategies, but also design computationally efficient near-optimal strategies. Moreover, the achievable performance of the system is approximately assessed.

Another dynamic resource allocation method, namely transceiver optimization is presented in Chapter 5. We consider a cognitive downlink transmission with one multi-antenna secondary base station (BS) and multiple mobile SUs. Robust transceiver filters are designed considering both of the following CSI error models: the bounded CSI error and the stochastic CSI error model. Due to the non-convexity of the problem, we resort to several distinct approaches for the development of efficient algorithms, such as alternating methods, the constrained gradient projection method, and most importantly, a method exploiting the uplink-downlink duality in the network. Furthermore, the performance of different algorithms is quantitatively compared.

Up to this point, spectrum sensing and interference mitigation strategies are investigated for a single paradigm. In Chapter 6, we consider a hybrid paradigm combining the interweave and the underlay paradigm. Specifically, the design of the power

allocation strategies in Chapter 4 is extended to additionally incorporate the reliability of the sensing outcome. Both soft-decision and hard-decision sensing results are explored. Moreover, we consider the location uncertainty of the primary network and provide a model of interference caused by the secondary transmission. The interference modeling result is also integrated into the power allocation problem. To this end, we exemplify the performance gain from "cognition" in CR systems by assessing the achievable performance of the SUs through the effective utilization of sensing observations.

Finally, Chapter 7 concludes the thesis and provides an outlook on potential future work.

Chapter 2

Cognitive Radio Systems

The emergence of cognitive radio has the potential to increase the spectral efficiency in traditional wireless communication systems where an exclusive spectrum usage is granted among different kind of users. Figure 2.1 exemplifies a network model in which both the PUs and the SUs coexist. The fundamental requirement for the design of the SUs' transceiver strategies is to limit the disturbance to primary systems. Based on different a priori information regarding the environment, cognitive radio systems mainly employ three paradigms for the DSA: the interweave, the underlay, or the overlay paradigm. Different approaches are used in different paradigms for the interference management, among which we focus on the two main perspectives. On the one hand, spectrum sensing, which characterizes the spectrum usage and the PUs' activity in the time, space, and frequency domains, can be considered as the basic learning function of the SUs and it is widely utilized in the interweave paradigm. Based on such information, the SUs adapt their transmission strategy, e.g., suspending the transmission or lowering the transmit power to avoid the interference to the PUs. On the other hand, dynamic resource allocation, e.g., power allocation and transceiver optimization, can effectively mitigate the interference when concurrent transmission of the PUs and the SUs is allowed. It can be realized by exploiting the available side information or the DoFs offered by the multi-user multiple-antenna systems.

In this chapter, we first briefly introduce three CR paradigms in Section 2.1. Two main techniques of the interference management applicable to CR systems are revised: spectrum sensing in Section 2.2 and dynamic resource allocation in Section 2.3. These two approaches provide the basis of the remainder the thesis.

2.1 Network Paradigms

There are three main paradigms of the CR system: the *interweave*, the *underlay*, and the *overlay* paradigm [38]. These paradigms are distinguished by the approaches to cooperate with the PUs. In the following, we recapitulate each of them, including the discussion on the corresponding requirements and the key techniques for interference management.

2.1.1 Interweave Paradigm

The interweave paradigm, also named *opportunistic spectrum access* (OSA), represents the opportunistic transmission of the SUs based on a priori knowledge of the available radio resource assigned to the PUs. The idea is based on the original motivation of

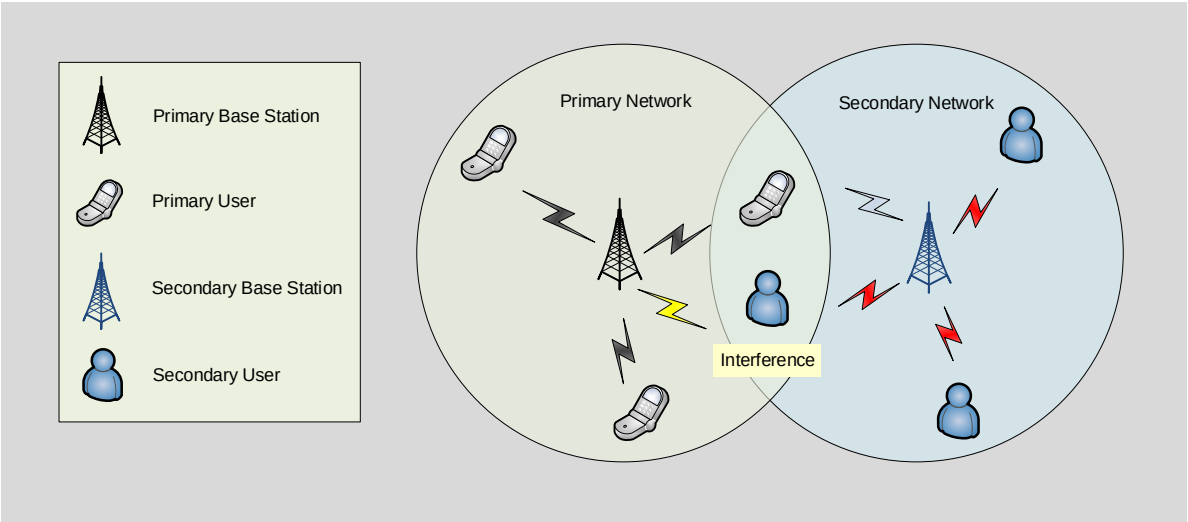

Figure 2.1: Coexistence of a primary and a secondary infrastructure-based network.

CR that the spectrum resource is under-utilized as shown by various measurement surveys, e.g., in [1,134]. Hence, spectral efficiency is improved by allowing the SUs to opportunistically access the resource when it is unoccupied by the PUs.

The preliminary for the SUs is to obtain the activity information of the PUs in space, time, and frequency [125]. To enable this requirement, the SUs can make use of the geo-location databases, beacon signals, or spectrum sensing [139]. Among these strategies, spectrum sensing embodies mostly the "cognitive" concept that the CRs need to intelligently learn the available side information from the environment. Moreover, spectrum sensing requires minimal infrastructure changes of the PUs' systems. Therefore, we concentrate on spectrum sensing methods in this paradigm. Through periodical monitoring and detecting of the occupancy status of the spectrum resources, the SUs transmit over the vacant resources, without yielding severe performance degradation to the PUs.

Although the original understanding of the spectrum sensing is the characterization of the spectrum opportunities, its general meaning also involves determining other characteristics of the spectrum usage, such as user activity and channel statistics, modulation schemes of the signal, carrier frequency, and bandwidth, etc. Basically, spectrum sensing exemplifies the typical learning functionality of CR systems.

2.1.2 Underlay Paradigm

The underlay paradigm, also entitled *spectrum sharing*, permits the concurrent transmission of primary and cognitive users as long as the interference at the PUs is properly limited. In this setting, the SUs need to have the knowledge of the potential performance degradation caused to the PUs, e.g., the interference power at the primary receivers or the outage probability of the primary transmissions. In order to control the interference, the initial idea is based on techniques of spread spectrum

and ultra-wideband (UWB) communication in which the signals from the SUs are spread below the noise level. However, such techniques may lead to conservative transmit strategies of the SUs due to very low transmit powers. Alternatively, several interference mitigation techniques, such as adaptive power allocation or transceiver optimization, allow for a control on the interference by exploiting the available CSI or multiple antennas at the SUs. Therefore, interference mitigation has raised a lot of research interest in this paradigm.

2.1.3 Overlay Paradigm

Similar to the underlay paradigm, the overlay paradigm also allows for the concurrent transmission of primary and cognitive users. The difference w.r.t. the underlay paradigm is that the interference to the primary transmission is compensated by using part of the secondary transmission to relay the primary message. To make it feasible, the underlying information required at the SUs is the knowledge of the PUs' codebooks, messages, and the channel gains. With such information, the SUs are able to cancel or mitigate the interference, thus, enhance spectral efficiency of the network. However, the practical realization is very challenging since it demands the SUs to have the most a priori PUs' knowledge among the three paradigms, especially the codebook and messages of one specific user type might be confidential for other kinds of users.

To summarize, the interweave paradigm aims at guaranteeing the exclusive transmission of the PUs and the SUs, whereas the underlay and the overlay paradigms mandate the simultaneous transmission of both kinds of users. Due to the practical difficulty in realizing the overlay paradigm, we focus on the investigation of the first two paradigms. Moreover, hybrid paradigms combining parts of the three paradigms are also possible. For instance, the works [102, 107, 117] showed that an enhanced performance is achieved by the SUs when jointly considering the interweave and the underlay paradigm.

Since the ultimate requirement for the cognitive transmission is to ensure that the performance degradation to the PUs is properly controlled, we outline different interference management strategies used in different paradigms. Specifically, the interweave paradigm requires the activity information of the PUs which can be obtained through spectrum sensing, whereas the underlay and overlay paradigms require proper interference mitigation and cancelation approaches.

2.2 Spectrum Sensing

Spectrum sensing refers to the task of characterizing the spectrum usage and the activity information of the PUs. Incorporating noise uncertainty, multi-path fading, and the shadowing effect into the specification, the SUs are required to quickly identify the presence of the primary signal in the very low signal-to-noise ratio (SNR) region [113, 124]. In the IEEE 802.22 WRAN standard, the SUs should be able to

sense the incumbent TV broadcasting transmission for power as low as $-116\,\mathrm{dBm}$, and wireless microphone transmission for power at $-107\,\mathrm{dBm}$ in less than 2 seconds [3, 118]. Such requirements regarding the sensitivity, agility, and reliability of the sensing are challenging for local spectrum sensing techniques. Furthermore, the hidden primary user problem might degrade the sensing performance by a single SU [139]. For example, if a SU is obstructed from the primary transmit signal, it might cause the unwanted interference to the PU due to a missed-detection of the PU's transmission. To solve the hidden primary user problem, one effective approach is to exploit more DoFs in the network, e.g., making a joint decision based on multiple sensing observations collected from multiple SUs. Hence, cooperative spectrum sensing is desirable to improve the sensing performance. Comprehensive surveys on spectrum sensing have been conducted in [11, 139, 142].

In this section, we outline some commonly-used spectrum sensing strategies. Beginning with local sensing methods, including the conventional energy detection and some feature detection methods, we proceed to the introduction of cooperative sensing which effectively ameliorates the local sensing subject to the hidden primary user problem and thus increases the sensing reliability.

2.2.1 Local Spectrum Sensing

Local spectrum sensing is formulated as a binary hypothesis test problem as

$$\mathbf{y}[n] = \begin{cases} \boldsymbol{\omega}[n], & \mathcal{H}_0 \\ \mathbf{x}[n] + \boldsymbol{\omega}[n], & \mathcal{H}_1 \end{cases} \tag{2.1}$$

with $\forall\, n = 0, \ldots N-1$. In eq. (2.1), $\mathbf{x}[n]$ represents the nth sample of the primary signal and $\mathbf{y}[n]$ indicates its received version at the SU. The noise term at the nth sample is $\boldsymbol{\omega}[n]$ which follows the proper complex Gaussian distribution. We assume that $\boldsymbol{\omega}[n] \sim \mathcal{CN}\left(\mathbf{0}, \sigma^2\mathbf{I}\right)$ holds with the noise variance σ^2. The dimensions of the vectors $\mathbf{x}[n]$, $\mathbf{y}[n]$, and $\boldsymbol{\omega}[n]$ are the same. We set the dimension equal to L, e.g., if there are L receive antennas at the SU. The observation interval includes N samples. The null hypothesis that the PU is absent is denoted by $\mathcal{H}_0$ and its presence is given by $\mathcal{H}_1$. Stacking the vectorial observation of different samples into one vector, we obtain the compact form of the received data at the SU

$$\mathbf{y} = \left[\mathbf{y}[1]^T, \ldots, \mathbf{y}[N]^T\right]^T. \tag{2.2}$$

The task of signal detection is to decide for the status $\mathcal{H}_0$ or $\mathcal{H}_1$. To achieve this goal, the test statistic $\Lambda(\mathbf{y})$ is formed based on the observation data $\mathbf{y}$ and then compared to a predefined threshold η [66]:

$$\Lambda(\mathbf{y}) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \eta. \tag{2.3}$$

The commonly used performance measures of the detector are the false alarm rate P_{FA} that indicates the probability to declare $\mathcal{H}_1$ given the actual status $\mathcal{H}_0$ and the probability of detection P_{D} that represents the probability to declare $\mathcal{H}_1$ given the status $\mathcal{H}_1$:[1]

$$P_{\mathrm{FA}} = \Pr\left\{\Lambda\left(\mathbf{y}\right) > \eta | \mathcal{H}_0\right\} \tag{2.4}$$

$$P_{\mathrm{D}} = \Pr\left\{\Lambda\left(\mathbf{y}\right) > \eta | \mathcal{H}_1\right\}. \tag{2.5}$$

In the IEEE 802.22 WRAN standard, the required P_{FA} is 10 % and the required P_{D} is 90 % [118]. For the purpose of visualizing the performance, the receiver operating characteristic (ROC) curve is used to show P_{D} as a function of P_{FA}. By choosing different values of η, different operating points, i.e., different pairs of P_{D} and P_{FA} values, are selected along the ROC curves.

The essential part in the design of spectrum sensing algorithms is to obtain the test statistic $\Lambda\left(\mathbf{y}\right)$ and the threshold η. According to the detection theory, they can be designed, e.g., under the Neyman-Pearson (NP) or the Bayesian criterion [66, 103]. Both result in the same form of the test statistic called the likelihood ratio

$$\Lambda\left(\mathbf{y}\right) = \frac{p\left(\mathbf{y}|\mathcal{H}_1\right)}{p\left(\mathbf{y}|\mathcal{H}_0\right)} \tag{2.6}$$

where $p\left(\mathbf{y}|\mathcal{H}_i\right)$ denotes the probability density function (PDF) of the observation $\mathbf{y}$ under the hypothesis $\mathcal{H}_i$, $i = 0, 1$. In the following, we introduce several widely used local spectrum sensing methods.

2.2.1.1 Energy Detection

One of the basic detection methods is known as energy detection (ED) [129]. Specifically, the received signal energy is measured over the observation interval. The accumulated metric is compared to a predefined threshold to decide whether the PUs' transmission is active or not. The resulting test statistic is

$$\Lambda_{\mathrm{ED}}\left(\mathbf{y}\right) = \frac{\|\mathbf{y}\|^2}{\sigma^2} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \eta. \tag{2.7}$$

The threshold η is set according to the system requirements, e.g., the fixed false alarm rate. If the primary signal is unknown to the detector and if signal samples are modeled as independent and identically distributed (i.i.d.) Gaussian random variables, the ED is the optimal detector according to the NP criterion.

The performance of the ED has been studied in [26] and sometimes can be expressed in closed form. For example, if the transmit signal vector $\mathbf{x}[n]$ contains binary phase-shift keying (BPSK) modulated signals with energy-per-symbol E_b, the test statistic $\Lambda_{\mathrm{ED}}\left(\mathbf{y}\right)$ follows a central chi-square distribution under the hypothesis

[1] For the simplification of illustration, herein we claim $\mathcal{H}_0$ for the special case that $\Lambda\left(\mathbf{y}\right) = \eta$. In practice, we can make equiprobable decisions of $\mathcal{H}_0$ and $\mathcal{H}_1$ for $\Lambda\left(\mathbf{y}\right) = \eta$. The performance measures can be deduced analogously.

$\mathcal{H}_0$ with degree of freedom $2LN$, while it follows a non-central chi-square distribution under the hypothesis $\mathcal{H}_1$ with degree of freedom $2LN$ and non-centrality parameter $2E_b/\sigma^2$. Closed-form expressions of P_{FA} and P_{D} can then be obtained using the cumulative density function (CDF) of the central or non-central chi-square distribution. Alternatively, if each element in $\mathbf{x}[n]$ is a complex Gaussian random variable with zero mean and variance σ_s^2, e.g., this assumption approximately holds if the primary signal is an orthogonal frequency-division multiplexing (OFDM) signal [79], the test statistic $\Lambda_{\mathrm{ED}}(\mathbf{y})$ under $\mathcal{H}_1$ becomes a scaled central chi-square distribution with degree of freedom $2LN$ and scaling factor $\left(\sigma_s^2+\sigma^2\right)/\sigma^2$.

The ED requires perfect knowledge of the noise variances. Due to the noise uncertainty in practical systems, the ED cannot make an accurate decision given that the SNR is below a certain threshold even for infinitely long sensing durations. Such an SNR threshold is called the SNR wall [124]. In fact, the uncertainty in the noise modeling causes a performance degradation in any moment-based detector. To alleviate this drawback, the work [124] proposed noise calibration schemes to enhance the noise modeling. Alternatively, the embedded features in the primary signals can be exploited to estimate the missing knowledge of the noise variances, rendering sensing algorithms robust to the noise uncertainty [148].

2.2.1.2 Matched Filter

In practical systems, pilots or preambles are usually transmitted periodically with the primary signals to help the receivers perform the channel estimation, synchronization, etc. If such information of the PUs is obtained at the SUs, e.g., provided by the primary systems' standard, it can be exploited to perform the matched filter (MF)-based sensing. In particular, if the primary signal vector $\mathbf{s} \in \mathbb{C}^{LN\times 1}$ is fixed and known to the SUs, the optimal approach according to the NP criterion for signal detection is MF with the test statistic

$$\Lambda_{\mathrm{MF}}(\mathbf{y}) = \mathcal{R}\left(\mathbf{y}^H\mathbf{s}\right). \tag{2.8}$$

The distribution of the test statistic for an additive white Gaussian noise (AWGN) channel can be analogously derived according to [66]. A example of its application in the IEEE 802.22 WRAN standard was shown in [23].

The advantage of MF-based detection mainly consists of three aspects. First, it is able to distinguish the primary signals from interferences and noises. Second, it works well even in the low SNR region. Third, compared to the ED, it requires a shorter observation time to achieve the predefined performance target. However, this coherent detection has its limitations. For instance, it requires the knowledge of the primary signal at both the physical layer (e.g., pulse shaping and modulation type of the pilot) and the medium access control layer (e.g., synchronization). Furthermore, the SUs are required to perform timing, carrier synchronization, and even equalization to detect a dedicated kind of primary signals, which increases the implementation complexity of the sensing method.

2.2.1.3 Cyclostationary-based Detection

The principle of cyclostationary detection is to exploit the embedded periodicity of the statistics of the man-made signals which may be caused by modulation and coding, the cyclic prefix, or intentionally inserted signals to aid the synchronization, channel estimation, or equalization. This kind of signals can usually be modeled as a cyclostationary random process. Defining the autocorrelation function of a discrete-time random process $y[n]$ as

$$r_y[n,\tau] = \mathbb{E}\{y[n]y^*[n+\tau]\} \tag{2.9}$$

the random process $y[n]$ is wide-sense cyclostationary if $r_y[n,\tau]$ is a periodic function in n for any τ [32]. Mathematically, this feature of the observed signal can be extracted by analyzing the cyclic autocorrelation in the time domain or its equivalent Fourier transformation called cyclic spectrum in the frequency domain

$$\textbf{cyclic autocorrelation}: \quad R_y^{\alpha}(\tau) = \frac{1}{N}\sum_{n=0}^{N-1} r_y[n,\tau]\, e^{-j2\pi\alpha n} \tag{2.10}$$

$$\textbf{cyclic spectrum}: \quad S_y^{\alpha}(f) = \sum_{\tau} R_y^{\alpha}(\tau) e^{-j2\pi f\tau} \tag{2.11}$$

where α denotes the cyclic frequency and N is the observation length. Basically, the test statistic of cyclostationary-based detection is a function of $R_y^{\alpha}(\tau)$ and $S_y^{\alpha}(f)$. For example, the cyclostationary-based feature detector using one cyclic frequency has been proposed in [25] and lately extended to use multiple cyclic frequencies in [86].

In general, the cyclic features of different types of signals are distinct. Thus, the main advantage of cyclostationary-based detection is its capability of differentiating between specific types of primary signals, interferences, and noises. Nevertheless, cyclostationary-based detection is vulnerable to model uncertainties and the channel fading effect [123]. Moreover, it is susceptible to sampling clock offsets [128].

2.2.1.4 Subspace-Based Detection

The idea of subspace-based detection is that the primary signal received at the SU is usually correlated, e.g., due to multiple receiver antennas or over-sampling effect. On the contrary, the noise at the SU is in general spatially and temporally white. This feature can be exploited in spectrum sensing to distinguish the cases containing the existence of the primary signals or pure noise, respectively. Due to the fact that such a feature is extracted from the specific structure of the sample covariance matrix, the corresponding algorithms are usually termed subspace-based detection methods.

In order to measure the signal whiteness, we choose some covariance based statistical parameters to characterize the subspace features. Specifically, considering the signal model in (2.1), the sample covariance matrix $\hat{\mathbf{R}}_y$ of the received signal is

$$\hat{\mathbf{R}}_y = \frac{1}{N}\sum_{n=0}^{N-1} \mathbf{y}[n]\mathbf{y}^H[n]. \tag{2.12}$$

Let $\lambda_i,\ \forall\, i = 0, \dots L-1$ be the eigenvalues of $\hat{\mathbf{R}}_y$ sorted in descending order of the magnitude. Several recent works designed the test statistics using these eigenvalues. For instance, maximum-minimum eigenvalue (MME) detection in [140] proposed the test statistic as the ratio of the maximum and minimum eigenvalues

$$\Lambda_{\mathrm{MME}}(\mathbf{y}) = \frac{\max_i \lambda_i}{\min_i \lambda_i} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \eta. \tag{2.13}$$

The distribution of $\Lambda_{\mathrm{MME}}(\mathbf{y})$ has been approximated using random matrix theory. Therefore, a decision threshold can be derived given that the false-alarm probability is fixed. Alternatively, the covariance based test statistic, e.g., the ratio of its diagonal and off-diagonal elements in $\hat{\mathbf{R}}_y$ was designed in [141]. The asymptotic distribution of the ratio was also derived for a selected decision threshold.

In a practical scenario, a priori knowledge of the primary signal space and the noise variance matrix might not be available at the SUs. In detection theory, the GLRT principle is a widely used approach to tackle the binary hypothesis testing problem with unknown parameters [66]. Asymptotically, the GLRT test statistic is equivalent to the uniformly most powerful (UMP) test among all invariant statistical tests [76]. The detailed introduction of the GLRT principle will be presented in Section 3.1. Recently, the GLRT framework has been applied to the context of developing the subspace-based spectrum sensing problem when only certain a priori information is available. For instance, The authors in [148] studied the case assuming spatially white noise with the same variance at each receiver and the unknown signal space of arbitrary rank, while the authors in [122, 132] considered the similar problem but assuming that the rank of the unknown signal space is equal to one. Considering the noise variances are different at each receiver and the rank of the unknown signal space is one, the sensing strategy was designed in [84]. Exploiting the information that the unknown OFDM signal covariance matrix has known eigenvalue multiplicities, the GLRT framework is applied in [9].

However, the potential utilization of the rank information has not been fully addressed. The rank information which indicates the number of independent signal sources can be estimated by the methods in [75, 133] or provided by the PU standard. For example, if sensing channels are mutually independent, the rank information is the number of independent signal streams transmitted by the PU. The rank information is useful to extract the signal space embedded in the sample covariance matrix of the received signals, especially for the case of a rank-deficient signal space. For example, a sensing method with rank information of the primary signal under the assumption of equal noise variances at the SUs was studied in [16]. However, the sensing method in a more general scenario where the noise variances are distinct at multiple SUs is not clear. This setting leads to the difficulty in deriving a GLRT-based method since the closed-form maximum likelihood estimator (MLE) of the unknown parameters is mathematically intractable to the best of our knowledge.

In summary, subspace-based approaches exploit the correlated nature of the primary signals which is in contrast to the whiteness of the noises. This feature is gen-

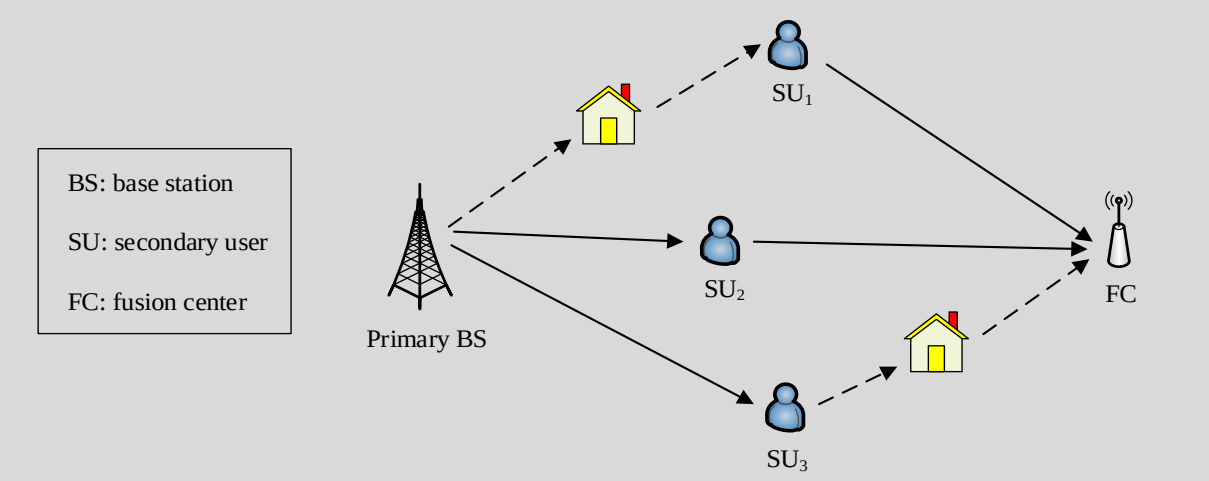

Figure 2.2: Hidden primary user problem in spectrum sensing.

erally embedded inside the primary signals, or can also be artificially constructed, e.g., using multiple receive antennas at the SUs. Under the circumstance of limited knowledge regarding the primary signal space and the noise variances, the GLRT principle can be applied to the subspace-based sensing methods to effectively tackle the problem.

2.2.2 Cooperative Spectrum Sensing

The local sensing by a single SU is subject to multi-path fading and shadowing. Fig. 2.2 depicts one example called the hidden primary user problem: the link from the primary BS to the first SU is in a deep fading, i.e., obstructed by a large obstacle from the primary transmitter (PT). Consequently, it might cause unexpected interference to the PU since the received primary signal power is too weak to detect. This uncertainty can be effectively mitigated by the spatial diversity in cooperative sensing. Additionally, cooperative sensing can relax the sensitivity requirement of local sensing and increase the agility of making a sensing decision.

The cooperative schemes are classified according to the cooperation manner: distributed manner and centralized manner. On the one hand, in centralized sensing, there is a central unit existing in the network to collect the sensing observations from the SUs and make a common sensing decision. This sensing manner can be widely applied to the infrastructure-based secondary network where a cognitive BS or access point (AP) serves as the fusion center (FC) to collect the sensing decisions from cooperating SUs. On the other hand, in distributed sensing, the SUs share the information among each other and aim at reaching a final sensing decision. It is usually applied in distributed networks, e.g., mobile ad hoc networks, where the users might explore consensus-based methods [78]. Compared to the centralized manner, although a distributed spectrum sensing system does not require the existence of the FC, an

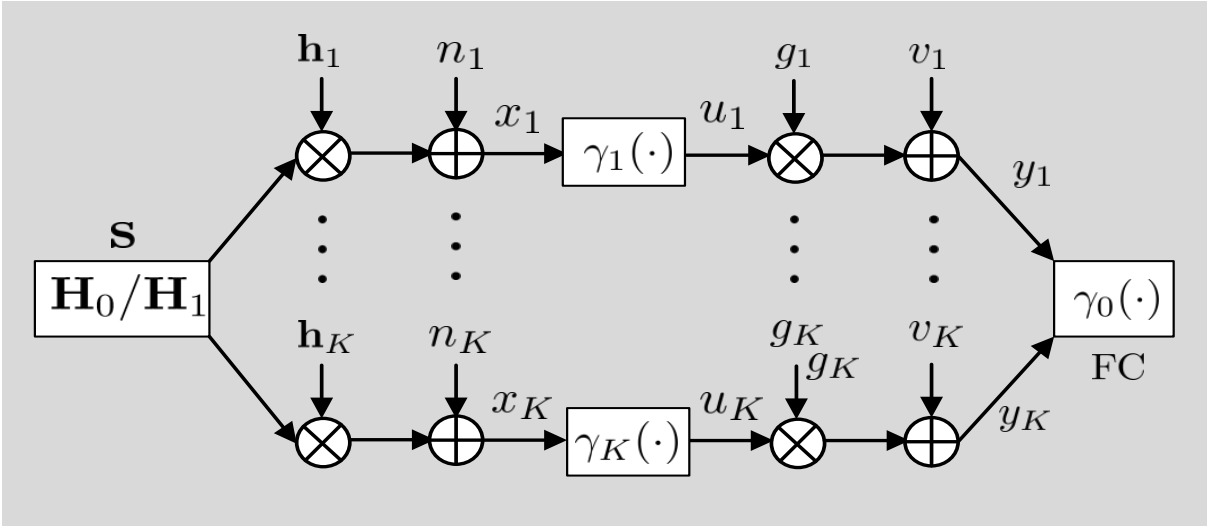

Figure 2.3: Fusion-based cooperative sensing.

increased implementation complexity is required at each SU and overhead channels are still needed to coordinate all the SUs. Therefore, we focus on the study of the centralized cooperative sensing.

The structure of cooperative sensing is illustrated in Figure 2.3 assuming there are K SUs in the network. Under the circumstance $\mathcal{H}_0$ or $\mathcal{H}_1$, the primary signal $\mathbf{s}$ is transmitted through the independent fading channels $\mathbf{h}_1, \ldots, \mathbf{h}_K$ to each user. The additive Gaussian noises of the sensing and reporting channels are represented by n_k and v_k, $\forall\, k = 1, \ldots, K$, respectively. Based on the local decision rule $\gamma_k(\cdot), \forall\, k = 1, \ldots, K$, the individual user k forwards its sensing decision $u_k = \gamma_k(x_k)$ to the FC through the reporting channel g_k and the noise v_k. The FC accumulates the noisy decisions y_k, $\forall\, k = 1, \ldots, K$, from all the sensors and finally makes the cooperative sensing decision based on the decision rule $\gamma_0(\cdot)$.

Depending on the type of information collected at the secondary FC, cooperative sensing can be categorized into soft-decision fusion and hard-decision fusion.

- *Soft-decision fusion*: For the soft-decision fusion, each SU sends soft values, e.g., received sensing data or energy, to the FC without quantization. Assuming the received signals from K cooperative SUs are mutually independent given $\mathcal{H}_0$ or $\mathcal{H}_1$, the log likelihood ratio (LRT) at the FC is

$$\begin{aligned}\Lambda^0_{\mathrm{LRT}} &= \ln\left(\frac{p\left(y_1, \ldots, y_K|\mathcal{H}_1\right)}{p\left(y_1, \ldots, y_K|\mathcal{H}_0\right)}\right) \\ &= \ln\left(\prod_{k=1}^{K}\frac{p\left(y_k|\mathcal{H}_1\right)}{p\left(y_k|\mathcal{H}_0\right)}\right) \\ &= \sum_{k=1}^{K}\ln\left(\frac{p\left(y_k|\mathcal{H}_1\right)}{p\left(y_k|\mathcal{H}_0\right)}\right)\end{aligned}$$

$$= \sum_{k=1}^{K} \Lambda_{\mathrm{LRT}}^{k} \tag{2.14}$$

where $\Lambda_{\mathrm{LLR}}^{0}$ is the LRT test statistic at the FC and $\Lambda_{\mathrm{LLR}}^{k}$, $\forall\, k = 1, \ldots, K$, is the LRT test statistic at the kth SU. Eq. (2.14) indicates that the optimal fusion rule is the sum of the local LRTs from all cooperative SUs under the mutual independence assumption. Correlation among the secondary signals can degrade the cooperative sensing performance [36].

- *Hard-decision fusion*: In order to reduce the signaling overhead between the SUs and the FCs, each SU can transmit the individual sensing decision by binary or BPSK signaling to the FC. The FC combines the hard decisions into one common decision, e.g., by performing the classical AND, OR, or voting rules [77, 150]. This scheme is named hard-decision sensing.

In general, soft-decision fusion outperforms hard-decision fusion at the expense of a larger signaling overhead to forward the data. Alternatively, quantized decision fusion is desirable to achieve the tradeoff between these two issues.

2.3 Dynamic Resource Allocation

Due to the hierarchical usage of the resources by primary and secondary systems, the main objective in the design of secondary systems is to constrain the performance degradation caused to the PUs below some tolerable limit. Dynamic resource allocation is widely considered as an effective way to compromise interference mitigation and improving the performance of secondary systems, cf. [147]. To achieve this goal, the SUs adjust the transceiver strategies, such as transmit power, bandwidth, precoders, and equalizers, according to the channel variations and the available a priori information regarding the wireless communication environment. For instance, if the CSI of the secondary links is in a favorable status and the primary user is inactive, the SUs can "greedily" use the available resource for transmission, and vice versa. Most importantly, the optimization of the secondary transceiver parameters are constrained by the performance degradation caused to the PUs. A commonly used performance measure is the received interference power at the PR termed the IT [33], either in the long-term or short-term manner [146].

In this section, we concentrate on introducing two key techniques for dynamic resource allocation: power allocation and transceiver optimization. Their effectiveness in interference management has been extensively studied in conventional multiuser cellular networks. Herein, we focus on their application in the DSA setting.

2.3.1 Power Allocation

In spectrum sharing networks, the secondary transmitters (STs) need to properly adapt the power for the sake of enhancing the throughput of the secondary network

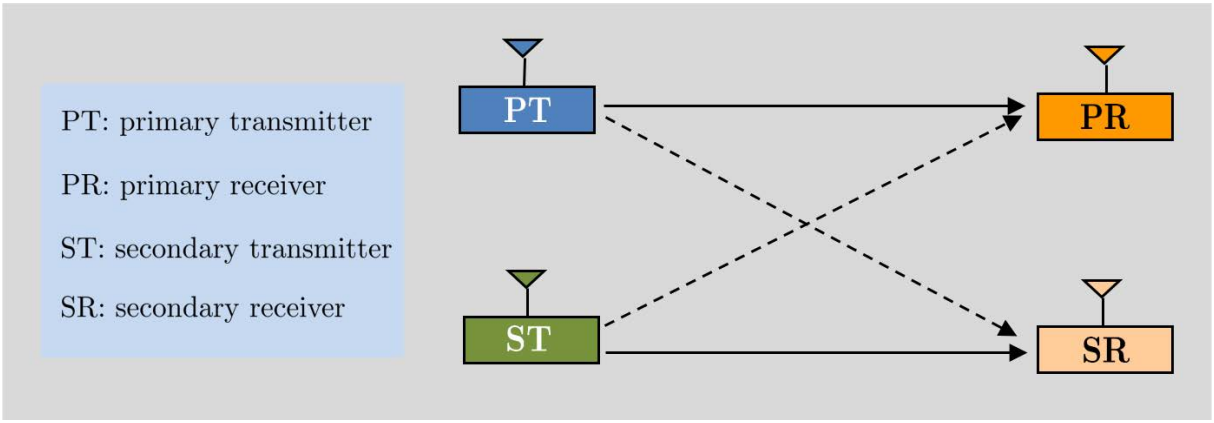

Figure 2.4: System model considered for power allocation.

and limiting the interference to the PUs. One example of the system model is depicted in Figure 2.4. A single secondary link coexists with a single primary link when each transmitter or receiver is equipped with a single antenna. The methodology in the power control design can also be applied to more sophisticated scenarios, e.g., with multiple PUs and SUs. In order to protect the primary transmission, various works consider an IT constraint in the design of secondary transmit strategies. For example, Gastpar studied the achievable rate of the secondary network over additive white Gaussian noise channels, subject to an average IT constraint [33]. In [37], Ghasemi and Sousa proposed the optimal power allocation strategies to maximize the secondary achievable rate over different types of fading channels, considering either a peak or an average IT constraint. The work [64] studied the optimal power allocation when jointly imposing a transmit power and an IT constraint. Considering both power constraints and incorporating the spectrum sensing result, the work [8] designed adaptive power and rate allocation schemes, and the work [120] studied the outage capacity under truncated channel inversion with a fixed rate strategy. Alternatively, limiting the primary capacity loss is the ultimate goal in regulating the secondary transmission. To this end, Zhang investigated the optimal power control under the PU's ergodic capacity-loss constraint in [149] and Kang *et al.* designed optimal power allocation methods constrained by an outage probability of the primary link in [65]. Both [149] and [65] demonstrated that a gain on the secondary rate is achieved by further exploiting the primary CSI instead of considering solely the IT constraint. Inspired by such an observation, the authors in [81] studied the power allocation problem to maximize the achievable rate of the secondary link subject to outage probability constraints on both primary and secondary links, and an average transmit power constraint. However, the prerequisite of all the above works is that the ST has instantaneous CSI of the entire network. In practice, this is challenging due to the difficulty in the cooperation between the PUs and the SUs.

In reality, the SUs can only acquire partial CSI of the links to the PUs due to limited cooperation with the PUs. Two kinds of partial CSI were considered for power

allocation design in spectrum sharing systems: outdated CSI [73, 89, 94, 110, 121] and statistical CSI [12, 28, 80, 116]. On the one hand, assuming that the ST has only outdated CSI from the ST to the PR link, the achievable rate of the secondary link was analyzed under an average interference power constraint in [94] or both a peak and an average power constraint in [110]. Nevertheless, the interference from the PT to the secondary receiver (SR) was not considered in [94] and [110]. Later, power allocation in the cross-interfering spectrum sharing system was investigated under an average power constraint [121] with outdated CSI. The impact of outdated CSI on the achievable rate of the secondary link was also thoroughly investigated in [73, 89], where the correlation of outdated CSI and the actual CSI was exploited to enhance the achievable performance of the secondary link. It was found that the low correlation between the outdated CSI and the actual CSI yields a large capacity loss of the secondary transmission. On the other hand, statistical CSI is desirable since it changes on a time scale that is much larger than the channel coherence time. Such CSI can be obtained by exploiting some side information, e.g., location information [20, 61, 137]. Given such statistical CSI the performance loss of the PUs can be limited in an average or a probability-constrained manner. For example, the achievable rate of the secondary link was investigated under an interference outage constraint in [80] and a minimum signal-to-interference-plus-noise ratio (SINR) constraint of the PR [28]. However, the interference from the PT to the SR was not considered in the above work. Taking this interference into account and assuming that the ST has perfect CSI of this interfering link, the authors in [116] designed the optimal and suboptimal power allocation approaches. Different CSI assumptions were made in [12] which studied power control methods aiming at achieving the minimum required mean rate of the secondary link subject to either a peak, an average transmit power constraint, or an outage constraint. The authors assumed that the ST has instantaneous CSI of the ST-SR link and the ST-PR link, but only statistical CSI of the PT-PR and the PT-SR link. Nevertheless, it is still challenging for the ST to obtain instantaneous CSI of the ST-PR link, since it requires the PR to estimate the CSI and feed it back to the ST.

Furthermore, the relation between the achievable performance of the SUs and the performance degradation of the PUs is desirable given certain power control strategies. The corresponding performance analysis can provide the insights on the benefit of exploiting the DSA and help to design guidelines in selecting system parameters for power adaptation strategies. However, closed-form expression of performance is always mathematically intractable. The alternative way is to develop a near-optimal strategy and characterize its analytical performance. Thus, the result can be used as a good approximation of the optimal achievable performance.

2.3.2 Transceiver Optimization

Spatial DoFs are widely exploited by employing multiple-antenna systems to keep the interference below a tolerable limit and maintain a high spectrum utilization. For instance, in traditional downlink cognitive systems, joint transceiver optimization is performed by using precoding at the BS and linear equalization at each mobile

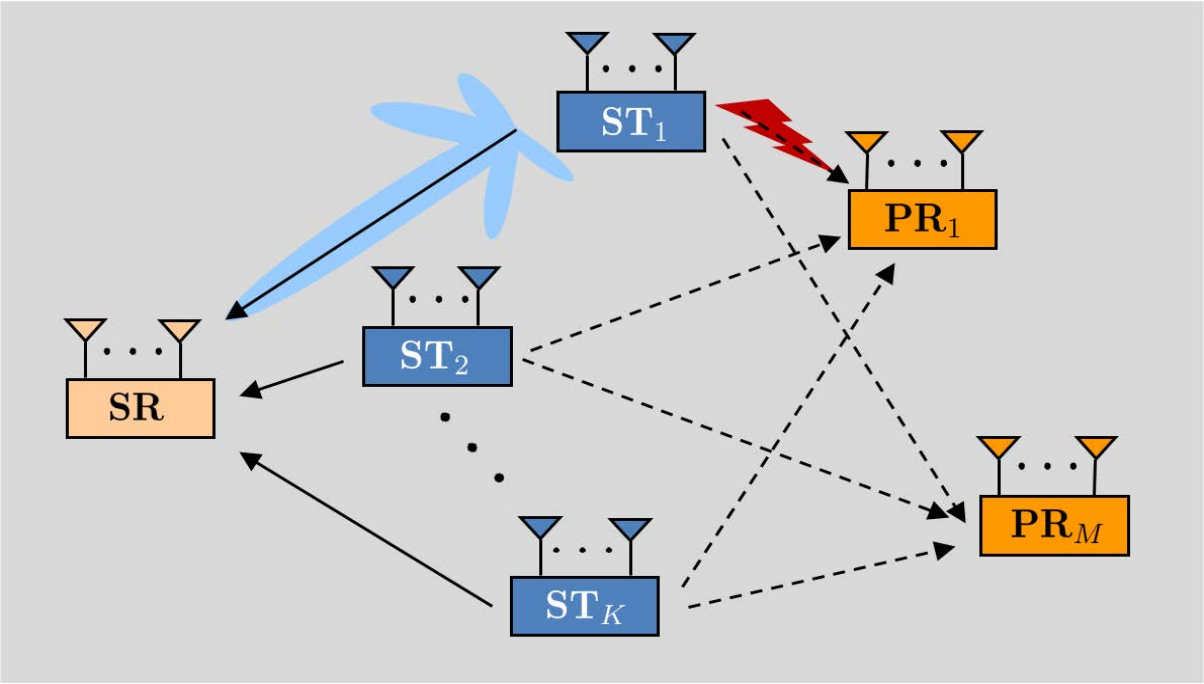

Figure 2.5: System model for transceiver optimization. Solid lines: desired links; dashed lines: interfering links.

terminal [114, 131, 135]. Specifically, a directional emission pattern is formed at the transmitter towards the desired user with the optimized precoders, while keeping the interference leaked to the undesired users low. Similarly, at the receiver side, interference and the effect of channel distortion are mitigated with the optimized equalizers. Recently, transceiver optimization techniques have been studied in CR systems, e.g., [60, 143]. Figure 2.5 exemplifies the scenario assuming a multiple access secondary network with K STs and one SR coexisting with M PRs. Contrary to omni-directional transmission, the first ST steers the transmit vector to the SR and avoids the transmission to the PRs. Consequently, the throughput of the secondary network is increased and the interference to the PRs is minimized.

Implementation of transceiver optimization requires CSI of the entire network at both the transmit and receiver sides. In reality, due to imperfect channel estimation, outdated estimates for fast time-varying channels or the quantization effect in the limited feedback, erroneous CSI is inevitable. Particularly in CR networks, the accurate channel knowledge is hard to acquire due to the limited cooperation with PUs. Since the transceiver optimization is quite sensitive to imperfect CSI [96], robust design becomes indispensable to deal with the channel uncertainties.

Two types of CSI error models are commonly considered: the bounded and the stochastic model. In general, the bounded model is applicable to frequency division duplex (FDD) systems, when the CSI error arises from the quantization of feedback information and is related to the quantization regions; the stochastic model applies to time division duplex (TDD) systems where the CSI is reused due to the reciprocity

of the uplink and downlink and its error is due to inaccurate channel estimation at the transmitter. Two principles are generally used for the system design with CSI errors. One is the worst-case approach to guarantee a certain system performance for any channel realizations within the CSI uncertainty region. For example, the authors in [144] studied the system composed of one cognitive user and one PU, where CSI of the ST-SR link is perfectly known but erroneous CSI of the ST-PR link is bounded within an ellipsoid uncertainty region. In [35], Gharavol et al. investigated the robust downlink beamforming design in multiuser multiple-input and single-output (MISO) cognitive networks, aiming at minimizing the transmit power of the secondary BSs, while targeting a lower bound on the received SINR of the SUs and imposing an upper limit on the interference power constraint at the PRs. Meanwhile the work in [153] considered the same scenario but aimed at maximizing the minimum of the received SINR of the SUs. The CSI uncertainty model was generalized from the bounded ball region to a bounded ellipsoid region. The other is the stochastic approach which guarantees a certain average system performance over all channel realizations within the uncertainty region [151].

In summary, transceiver optimization is an effective way for interference mitigation in CR networks. Due to practical difficulties, a robust design considering imperfect CSI is imperative.

Chapter 3

Spectrum Sensing with Limited A Priori Information

Spectrum sensing aims at detecting the presence or absence of the primary transmission. The main challenge is to detect the PUs' activity in the low SNR region in which the environment uncertainties, e.g., fading channels and noise uncertainty, can severely degrade the performance. Moreover, due to the limited cooperation between the PUs and the SUs, the information regarding the primary signals and the CSI of the entire network is usually not available at the SUs. Driven by the aforementioned issues, the design of efficient sensing strategies which are robust to the uncertainties and can fully exploit the available but limited a priori knowledge is necessary.

Recently, subspace-based spectrum sensing methods have raised a lot of research interest due to its effectiveness over the conventional ED approach and robustness to the environment uncertainties [16, 142]. However, these methods require the SUs to obtain complete a priori knowledge of the primary signal space and noise variances, which is unrealistic in practice. In detection theory, the GLRT principle is a widely used approach to deal with binary hypothesis testing problems with unknown parameters [66]. Since spectrum sensing is usually formulated as a binary hypothesis testing problem, the GLRT framework has recently been applied to solve the sensing problem when only certain a priori information is available, e.g., in [9,84,122,132,148].

In this chapter, we investigate cooperative spectrum sensing algorithms with limited a priori information. In order to provide a theoretical basis for later use, a brief review of the GLRT principle is first provided in Section 3.1. Then, two detailed sensing problems are studied: First, in Section 3.2, we exploit the rank information to efficiently extract the primary signal space and design the corresponding subspace-based sensing algorithm. Herein, the signals through reporting channels are assumed to be perfectly recovered. Second, in Section 3.3, we consider that the signals through sensing and reporting channels are subject to fading effects. Moreover, the structure of the primary signal spaces are unknown at the SUs. The GLRT-based sensing algorithms are derived by exploiting the available partial CSI of the channels.

The results presented in this chapter have been published in part by the author in [39, 45, 46][1].

[1] In reference to IEEE copyright material which is used with permission in this thesis, the IEEE does not endorse any of RWTH Aachen University's products or services. Internal or personal use of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution, please go to `http://www.ieee.org/publications_standards/publications/rights/rights_link.html` to learn how to obtain a License from RightsLink.

3.1 GLRT Principle

We consider a binary hypothesis testing problem with unknown parameter θ. The observation data $\mathbf{x}$ is used to make a choice between two competing hypotheses: the null hypothesis $\mathcal{H}_0$ and the alternative hypothesis $\mathcal{H}_1$. We denote θ_0 and θ_1 as the values of the unknown parameter θ under $\mathcal{H}_0$ and $\mathcal{H}_1$, respectively. The MLEs of θ_0 and θ_1 given as $\hat{\theta}_0$ and $\hat{\theta}_1$ are

$$\hat{\theta}_0 = \arg\max_{\theta_0} p(\mathbf{x}|\mathcal{H}_0, \theta_0) \tag{3.1}$$

$$\hat{\theta}_1 = \arg\max_{\theta_1} p(\mathbf{x}|\mathcal{H}_1, \theta_1). \tag{3.2}$$

GLRT-based methods rely on the conventional LRT method by replacing the exact knowledge of θ with the MLEs $\hat{\theta}_0$ and $\hat{\theta}_1$. Thus, the resulting test statistic is

$$\Lambda_{\text{GLRT}}(\mathbf{x}) = \ln \frac{p(\mathbf{x}|\mathcal{H}_1, \theta_1 = \hat{\theta}_1)}{p(\mathbf{x}|\mathcal{H}_0, \theta_0 = \hat{\theta}_0)}. \tag{3.3}$$

Consequently, the solution of the binary hypothesis problem is determined as

$$\Lambda_{\text{GLRT}}(\mathbf{x}) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \gamma_{\text{GLRT}} \tag{3.4}$$

where γ_{GLRT} is the pre-defined threshold. The decision in (3.4) indicates that we simply determine the outcome with the largest probability given the observation $\mathbf{x}$.

3.2 Spectrum Sensing with Rank Information

In a practical cognitive radio system, the exact knowledge of the primary signal space and noise variances is hard to obtain. Although several works investigated this kind of spectrum sensing problems [9, 84, 122, 132, 148], the potential utilization of the rank information of the primary signal space has not been fully addressed. The rank information indicates the number of independent signal sources. For example, if the channels of all sensing links are mutually independent, the rank information is the number of independent signal streams transmitted by the PUs. Such information can be estimated by the methods in [75, 133] or provided by the standard of primary systems. In what follows, we attempt to show how to use the rank information to extract and identify the signal space embedded in the sample covariance matrix of the received signals. With the estimated signal space at hand, GLRT-based sensing methods are then designed.

We consider a cooperative CR network with M SUs. Each SU is equipped with a single receive antenna and performs spectrum sensing based on the observation over N sampling times. Denoting the stacked receive signals of M SUs at the nth time

instant as $\mathbf{x}[n] \in \mathbb{C}^{M\times 1}$, $\forall\, n = 1, \ldots, N$, the task is to differentiate binary hypotheses where the absence and the presence of the PU are given by $\mathcal{H}_0$ and $\mathcal{H}_1$, respectively

$$\begin{aligned}\mathcal{H}_0 \ &: \mathbf{x}[n] = \mathbf{w}[n] \\ \mathcal{H}_1 \ &: \mathbf{x}[n] = \mathbf{H}\mathbf{q}[n] + \mathbf{w}[n]\end{aligned}$$

where the primary signal at the nth time instant is denoted by $\mathbf{q}[n] \in \mathbb{C}^{q\times 1}$. The CSI of the flat fading sensing channels is $\mathbf{H} \in \mathbb{C}^{M\times q}$. The received signal $\mathbf{H}\mathbf{q}[n]$ is temporally white and $\mathbf{H}\mathbf{q}[n] \sim \mathcal{CN}(\mathbf{0}, \mathbf{R}_S)$ follows, where $\mathbf{R}_S$ indicates its covariance matrix. The spatially and temporally white noises received at the SUs are given by $\mathbf{w}[n] \in \mathbb{C}^{M\times 1}$ and $\mathbf{w}[n] \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Lambda})$ with

$$\mathbf{\Lambda} = \mathrm{diag}\{\sigma_1^2, \ldots, \sigma_M^2\} \tag{3.5}$$

where the noise variance at the nth SU is given by σ_m^2, $\forall\, m = 1, \ldots, M$. We consider a general case that $\mathbf{\Lambda}$ has different diagonal entries due to the random layout of cooperative SUs in the network.

Defining $\mathbf{x} = \left[\mathbf{x}^T[1], \ldots, \mathbf{x}^T[N]\right]^T$, its PDF under $\mathcal{H}_i$ is given as

$$p(\mathbf{x}|\mathcal{H}_i) = \frac{1}{\pi^{MN}\left(\det \mathbf{\Sigma}_i\right)^N} \exp\left(-N\mathrm{tr}\left\{\mathbf{\Sigma}_i^{-1}\mathbf{R}_x\right\}\right) \tag{3.6}$$

where $\mathbf{\Sigma}_i$ is represented the covariance matrices of $\mathbf{x}$ under both hypotheses. Thus, we have

$$\mathbf{\Sigma}_0 = \mathbf{\Lambda} \tag{3.7}$$

$$\mathbf{\Sigma}_1 = \mathbf{R}_S + \mathbf{\Lambda} \tag{3.8}$$

and the sample covariance matrix of $\mathbf{x}$ is

$$\mathbf{R}_x = \frac{1}{N}\sum_{n=1}^{N} \mathbf{x}[n]\mathbf{x}^H[n]. \tag{3.9}$$

In practice, the noise variance can be calibrated at each SU. However, due to the noise uncertainty effect [124], the calibration error would deteriorate the sensing performance. In the following, we apply the GLRT principle to both scenarios assuming that the noise covariance matrix is known and unknown. In both scenarios, the structure of primary signal space is unknown at the SUs. Nevertheless, its rank information

$$\mathrm{rank}(\mathbf{R}_S) = r, \quad 0 \le r \le M \tag{3.10}$$

is known at the SUs.

3.2.1 Unknown Noise Variances

In this case, both covariance matrices $\mathbf{\Sigma}_i$, $i = 0, 1$ of the received signal $\mathbf{x}$ under $\mathcal{H}_0$ and $\mathcal{H}_1$ are unknown. Therefore, the main work in applying the GLRT principle is to calculate the MLEs of the unknown parameters.

Using the PDF expression in (3.6), the log-likelihood function (LLF) of $\mathbf{x}$ under the hypothesis $\mathcal{H}_0$ is

$$\ln p(\mathbf{x}|\mathcal{H}_0, \mathbf{\Lambda}) = -MN \ln \pi - N \ln \det \mathbf{\Lambda} - N\text{tr}\{\mathbf{\Lambda}^{-1}\mathbf{R}_x\}. \tag{3.11}$$

where we use (3.7) since here the received signal $\mathbf{x}$ is only composed of the noise term. Applying (3.1), the MLE of $\mathbf{\Lambda}$ is obtained as

$$\hat{\mathbf{\Lambda}}|\mathcal{H}_0 = \text{diag}\{\mathbf{R}_x\}. \tag{3.12}$$

Similarly, under $\mathcal{H}_1$, the LLF of $\mathbf{x}$ is

$$\ln p(\mathbf{x}|\mathcal{H}_1, \mathbf{R}_S, \mathbf{\Lambda}) = -MN \ln \pi - N \ln \det(\mathbf{R}_S + \mathbf{\Lambda}) - N\text{tr}\left\{(\mathbf{R}_S + \mathbf{\Lambda})^{-1}\mathbf{R}_x\right\}. \tag{3.13}$$

Using

$$\bar{\mathbf{R}}_S = \mathbf{\Lambda}^{-1/2}\mathbf{R}_S\mathbf{\Lambda}^{-1/2} \tag{3.14}$$

$$\bar{\mathbf{R}}_x = \mathbf{\Lambda}^{-1/2}\mathbf{R}_x\mathbf{\Lambda}^{-1/2} \tag{3.15}$$

Eq. (3.13) is reformulated to

$$\begin{aligned}\ln p(\mathbf{x}|\mathcal{H}_1, \bar{\mathbf{R}}_S, \mathbf{\Lambda}) = &- MN \ln \pi - N \ln \det(\mathbf{I}_M + \bar{\mathbf{R}}_S) \\ &- N \ln \det \mathbf{\Lambda} - N\text{tr}\left\{(\mathbf{I}_M + \bar{\mathbf{R}}_S)^{-1}\bar{\mathbf{R}}_x\right\}.\end{aligned} \tag{3.16}$$

Note that the rank of $\bar{\mathbf{R}}_S$ is equal to that of $\mathbf{R}_S$.

To apply the GLRT principle, we need to calculate the MLE of $\bar{\mathbf{R}}_S$. This can be obtained by maximizing the LLF in (3.16). The result is given in the following lemma.

Lemma 3.2.1. *Consider the optimization problem*

$$\begin{aligned}\max_{\bar{\mathbf{R}}_S} \quad & -\ln \det(\bar{\mathbf{R}}_S + \mathbf{I}_M) - \text{tr}\left\{(\bar{\mathbf{R}}_S + \mathbf{I}_M)^{-1}\bar{\mathbf{R}}_x\right\} \\ s.t. \quad & \bar{\mathbf{R}}_S \succeq 0\end{aligned} \tag{3.17}$$

where the rank of $\bar{\mathbf{R}}_S$ is r. The optimal solution of (3.17) is denoted by $\widehat{\bar{\mathbf{R}}}_S$ and given as

$$\widehat{\bar{\mathbf{R}}}_S = \sum_{i=1}^{m_1} (\bar{\lambda}_{x,i} - 1)^+ \bar{\mathbf{u}}_{x,i}\bar{\mathbf{u}}_{x,i}^H \tag{3.18}$$

where $\bar{\lambda}_{x,i}$, $\forall\, i = 1, \ldots, M$ are the eigenvalues of $\bar{\mathbf{R}}_x$ in descending order of the magnitude and $\bar{\mathbf{u}}_{x,i}$ are the corresponding eigenvectors. Assuming r_1 is the largest index of $\bar{\lambda}_{x,i}$ which satisfies $\bar{\lambda}_{x,i} \geq 1$, then we have $m_1 = \min(r, r_1)$.

Proof. We first rewrite the two terms in the objective function of (3.17) which contain $\bar{\mathbf{R}}_S$ as

$$\det(\bar{\mathbf{R}}_S + \mathbf{I}_M) = \prod_{i=1}^{r} (\bar{\lambda}_{s,i} + 1) \tag{3.19}$$

$$(\bar{\mathbf{R}}_S + \mathbf{I}_M)^{-1} = \bar{\mathbf{U}}_S (\bar{\mathbf{\Lambda}}_S + \mathbf{I}_M)^{-1} \bar{\mathbf{U}}_S^H \tag{3.20}$$

where $\bar{\mathbf{U}}_S = [\bar{\mathbf{u}}_{s,1}, \ldots, \bar{\mathbf{u}}_{s,M}]$ is a unitary matrix containing eigenvectors of $\bar{\mathbf{R}}_S$. The diagonal matrix $\bar{\mathbf{\Lambda}}_S$ contains the eigenvalues of $\bar{\mathbf{R}}_S$ in descending order:

$$\bar{\mathbf{\Lambda}}_S = \operatorname{diag}\left\{\bar{\lambda}_{s,1}, \bar{\lambda}_{s,2}, \ldots, \bar{\lambda}_{s,r}, 0, \ldots, 0\right\}. \tag{3.21}$$

Considering the positive semidefiniteness (PSD) of matrix $\bar{\mathbf{\Lambda}}_S$ and its rank is equal to r, it implicitely requires that

$$\bar{\lambda}_{s,i} > 0, \quad \forall\, i = 1, \ldots, r. \tag{3.22}$$

Taking (3.19) and (3.20) into the objective function of (3.17) yields

$$\widehat{\bar{\mathbf{R}}}_S = \arg \min_{\bar{\mathbf{U}}_S, \bar{\mathbf{\Lambda}}_S \succeq 0} \; f\left(\bar{\mathbf{U}}_S, \bar{\mathbf{\Lambda}}_S\right) \tag{3.23}$$

with

$$f\left(\bar{\mathbf{U}}_S, \bar{\mathbf{\Lambda}}_S\right) = \sum_{i=1}^{r} \ln(\bar{\lambda}_{s,i} + 1) + \operatorname{tr}\left\{\bar{\mathbf{U}}_S (\bar{\mathbf{\Lambda}}_S + \mathbf{I}_M)^{-1} \bar{\mathbf{U}}_S^H \bar{\mathbf{R}}_x\right\}. \tag{3.24}$$

To solve the optimization problem (3.23), we first consider the optimization of matrix $\bar{\mathbf{U}}_S$ which contain M orthogonal normalized column vectors. The optimal $\bar{\mathbf{U}}_S$ to minimize the function $f\left(\bar{\mathbf{U}}_S, \bar{\mathbf{\Lambda}}_S\right)$ can be derived according to [7, Theorem 2] as

$$\bar{\mathbf{u}}_{s,i} = \bar{\mathbf{u}}_{x,i}, \quad \forall\, i = 1, \ldots, M. \tag{3.25}$$

Next, integrating (3.25) into $f\left(\bar{\mathbf{U}}_S, \bar{\mathbf{\Lambda}}_S\right)$ in (3.24), (3.24) is simplified as a function of $\bar{\mathbf{\Lambda}}_S$ only. The resulting new objective function $g\left(\bar{\mathbf{\Lambda}}_S\right)$ is

$$g\left(\bar{\mathbf{\Lambda}}_S\right) = \sum_{i=1}^{r} \left(\ln\left(\bar{\lambda}_{s,i} + 1\right) + \frac{\bar{\lambda}_{x,i}}{\bar{\lambda}_{s,i} + 1}\right). \tag{3.26}$$

Considering a new optimization variable

$$\tilde{\lambda}_{s,i} = \left(\bar{\lambda}_{s,i} + 1\right)^{-1}, \quad \forall\, i = 1, \ldots, r \tag{3.27}$$

we have the following property of $\tilde{\lambda}_{s,i}$ due to the positivity of $\bar{\lambda}_{s,i}$ in (3.22)

$$\tilde{\lambda}_{s,i} < 1, \quad \forall\, i = 1, \ldots, r. \tag{3.28}$$

Using (3.27), Eq. (3.26) is written as

$$\tilde{g}\left(\tilde{\mathbf{\Lambda}}_S\right) = \sum_{i=1}^{r} \left(-\ln \tilde{\lambda}_{s,i} + \bar{\lambda}_{x,i}\tilde{\lambda}_{s,i}\right) \tag{3.29}$$

where

$$\tilde{\mathbf{\Lambda}}_S = \operatorname{diag}\left\{\tilde{\lambda}_{s,1}, \tilde{\lambda}_{s,2}, \ldots, \tilde{\lambda}_{s,r}, 0, \ldots, 0\right\}. \tag{3.30}$$

It is easy to show that $\tilde{g}\left(\tilde{\mathbf{\Lambda}}_S\right)$ in (3.29) is a convex function of $\tilde{\lambda}_{s,i}, \forall\, i = 1, \ldots, r$. Taking the derivative of the above equation over each $\tilde{\lambda}_{s,i}$ and setting it to zero, we obtain the estimate of $\tilde{\lambda}_{s,i}$ as $\bar{\lambda}_{x,i}^{-1}$. Recalling the condition in (3.28), we have

$$\hat{\tilde{\lambda}}_{s,i} = \min\left(\bar{\lambda}_{x,i}^{-1},\ 1\right), \quad \forall\, i = 1, \ldots, r. \tag{3.31}$$

Consequently, the MLEs of $\bar{\lambda}_{s,i}$ is obtained from (3.27) and (3.31) as

$$\hat{\bar{\lambda}}_{s,i} = (\bar{\lambda}_{x,i} - 1)^{+}, \quad \forall\, i = 1, \ldots, r. \tag{3.32}$$

Combining (3.25) and (3.32) yields (3.18). This concludes the proof. □

Remark: The similar problem considered in Lemma 3.2.1 for $\mathbf{\Lambda} = \mathbf{I}$ was discussed in [16]. However, our result is different from that of [16] where m_1 is equal to r. We additionally consider the PSD characteristic of $\bar{\mathbf{R}}_S$ and choose $m_1 = \min(r, r_1)$.

The LLF in (3.16) is equal to the objective function in (3.17) plus a term that is independent of $\bar{\mathbf{R}}_S$. Therefore, the MLE of $\bar{\mathbf{R}}_S$ is given in (3.18). Integrating this result into (3.16), we get the LLF as a function of only one unknown parameter $\mathbf{\Lambda}$ since $\bar{\mathbf{R}}_x$ is a function of $\mathbf{\Lambda}$

$$\begin{aligned} \ln p(\mathbf{x}|\mathcal{H}_1, \mathbf{\Lambda}) = & -MN\ln\pi - N\ln\det\bar{\mathbf{R}}_x \\ & - N\sum_{i=m_1+1}^{M}\bar{\lambda}_{x,i} + N\ln\prod_{i=m_1+1}^{M}\bar{\lambda}_{x,i} - Nm_1. \end{aligned} \tag{3.33}$$

The optimal $\bar{\lambda}_{x,i}$ to maximize (3.33) is

$$\bar{\lambda}_{x,i} = 1, \quad \forall\, i = m_1 + 1, \ldots, M.$$

However, the optimal $\mathbf{\Lambda}$ to maximize (3.33) is hard to obtain [84] under this circumstance. Follow the idea in [84], we use the near-optimal MLE in the low SNR region of the received primary signals

$$\hat{\mathbf{\Lambda}}|\mathcal{H}_1 = \operatorname{diag}\{\mathbf{R}_x\}. \tag{3.34}$$

From Lemma 3.2.1, we see that $\hat{\bar{\mathbf{R}}}_S$ is a function of $\mathbf{\Lambda}$. However, the perfect information of $\mathbf{\Lambda}$ is not available in this case. Thus, we represent the MLE of $\bar{\mathbf{R}}_S$ by replacing $\mathbf{\Lambda}$ in $\hat{\bar{\mathbf{R}}}_S$ with its estimated value $\hat{\mathbf{\Lambda}}$. The resulting MLE is denoted by $\tilde{\bar{\mathbf{R}}}_S$ and given as

$$\tilde{\bar{\mathbf{R}}}_S = \sum_{i=1}^{m_2} (\tilde{\lambda}_{x,i} - 1)^+ \tilde{\mathbf{u}}_{x,i} \tilde{\mathbf{u}}_{x,i}^H \tag{3.35}$$

where $\tilde{\lambda}_{x,i}$, $\tilde{\mathbf{u}}_{x,i}$, $\forall\, i = 1, \ldots, M$ are the eigenvalues in descending order and the corresponding eigenvectors of the matrixes

$$\tilde{\bar{\mathbf{R}}}_x = \hat{\mathbf{\Lambda}}^{-1/2} \mathbf{R}_x \hat{\mathbf{\Lambda}}^{-1/2} \tag{3.36}$$

respectively. Assuming r_2 is the largest index of $\tilde{\lambda}_{x,i}$ which satisfies $\tilde{\lambda}_{x,i} \geq 1$, then $m_2 = \min(r, r_2)$ holds.

Finally, the test statistic of the first GLRT-based method called GLRT1 is calculated by inserting the MLEs into the general form (3.3):

$$T_{\text{GLRT1}} = \ln p(\mathbf{x}|\mathcal{H}_1) - \ln p(\mathbf{x}|\mathcal{H}_0) = N \left(\sum_{i=1}^{m_2} \tilde{\lambda}_{x,i} - \ln \prod_{i=1}^{m_2} \tilde{\lambda}_{x,i} - m_2 \right). \tag{3.37}$$

3.2.2 Known Noise Variances

In this case, the LLFs under both hypotheses are given in (3.11) and (3.13). The only difference to Section 3.2.1 is that the noise covariance matrix $\mathbf{\Lambda}$ is now known at the SUs. Thus, under $\mathcal{H}_0$, there is no unknown parameter; under $\mathcal{H}_1$, only the signal covariance matrix $\mathbf{R}_S$ is unknown. Its estimation under $\mathcal{H}_1$ is given in (3.18) expressed with the perfect noise information $\mathbf{\Lambda}$. Taking the MLE of $\mathbf{R}_S$ into the GLRT detector (3.3) results in the test statistic of the second GLRT-based method termed GLRT2

$$T_{\text{GLRT2}} = N \left(\sum_{i=1}^{m_1} \bar{\lambda}_{x,i} - \ln \prod_{i=1}^{m_1} \bar{\lambda}_{x,i} - m_1 \right). \tag{3.38}$$

where m_1 was given in Lemma. 3.2.1 that $m_1 = \min(r, r_1)$.

Remark: The results of the proposed GLRT-based methods include some special cases introduced in the literature.

- Considering that $\mathbf{\Lambda}$ has the same diagonal entries, i.e., the noise variances are the same for each SU. If $\mathbf{R}_S$ is full rank, the test statistic of GLRT1 reduces to the sphericity test in [92] with unknown $\mathbf{\Lambda}$, while the test statistic of GLRT2 reduces to the signal-subspace eigenvalue test in [148]. If the rank of the signal space is equal to one, both test statistics match the results in [132] and [122].

- Considering that $\mathbf{\Lambda}$ has different diagonal entries, i.e., the noise variances are not identical for each SU, and assuming $\mathbf{\Lambda}$ is unknown, if the rank of the signal space is one, the test statistic of GLRT1 is the same as the λ_1 test in [84].

Therefore, the proposed GLRT methods provide a general form of a binary hypothesis testing problem in which the noise covariance matrix is a diagonal matrix and the signal space is unknown but with available rank information.

3.2.3 Performance Evaluation

To evaluate the algorithm performance, we consider a CR network with $M = 10$ SUs detecting one PU over $N = 400$ samples. The SNRs at the SUs are equal to [-8, -10, -16, -12, -8, -10, -11, -14, -10, -12] dB, respectively. The noise uncertainty effect of the sensing links is evaluated. Specifically, the estimated noise variances at each SU are 1, 0.871, 1.15, 0.871, 1, 0.871, 1.15, 0.871, 0.871, and 1.15, respectively. The actual noise variance is $\tilde{\sigma}_m^2 = \sigma_m^2/\alpha$, where $10\log_{10}\alpha$ is uniformly distributed in the region $[-B, B]$ with the noise uncertainty factor B in dB [113]. For example, $B = 0$ dB indicates that the noise variance is precisely known.

Five sensing methods are compared:

- *ED*: The test statistic is the sum of the signal energy received at all SUs

$$T_{\text{ED}} = N\mathbf{R}_x. \tag{3.39}$$

- *GLRT1*: The test statistic is given in (3.37).
- *GLRT2*: The test statistic is given in (3.38).
- *Sphe.*: The sphericity test in [92] assuming the noise variances are the same for each link. The test statistic is

$$T_{\text{Sphe.}} = \frac{\frac{1}{M}\sum_{i=1}^{M}\bar{\lambda}_{x,i}}{\prod_{i=1}^{M}\left(\bar{\lambda}_{x,i}\right)^{1/M}}. \tag{3.40}$$

- λ_1 *test*: The test statistic can be similarly obtained as in [84] assuming the rank of the primary signal space is one

$$T_{\lambda_1} = \frac{\bar{\lambda}_{x,1}}{\frac{1}{M}\sum_{i=1}^{M}\bar{\lambda}_{x,i}}. \tag{3.41}$$

- *Rank ref.*: The closed-form test statistic in [108] assuming the rank of the primary signal space is known and noise variances are unknown

$$T_{\text{Rank ref.}} = \begin{cases} \dfrac{\prod_{i=1}^{r}[\mathbf{R}_x]_{i,i}}{\det\mathbf{R}_x}, & r \geq M - \sqrt{M} \\ \sum_{i=1}^{r}\left(\tilde{\lambda}_{x,i} - \ln\tilde{\lambda}_{x,i}\right), & \text{otherwise.} \end{cases} \tag{3.42}$$

The algorithm is also based on the GLRT principle[2].

[2] This work [108] is published after our conference paper [46] relevant to this section. Both works investigated the similar problem and shared some similarity in the derivation.

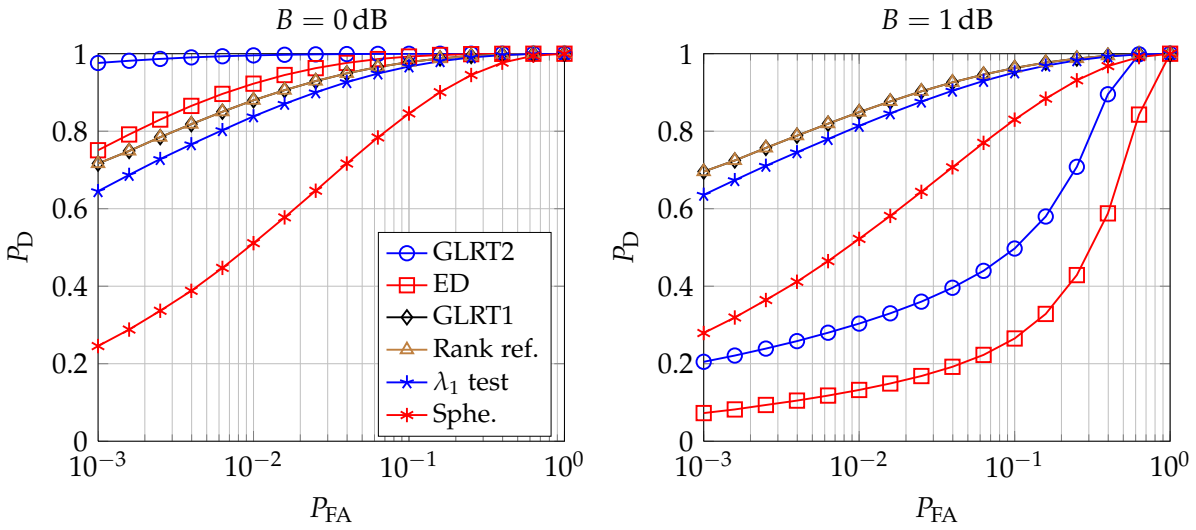

Figure 3.1: ROC curve for different noise uncertainty parameters B.

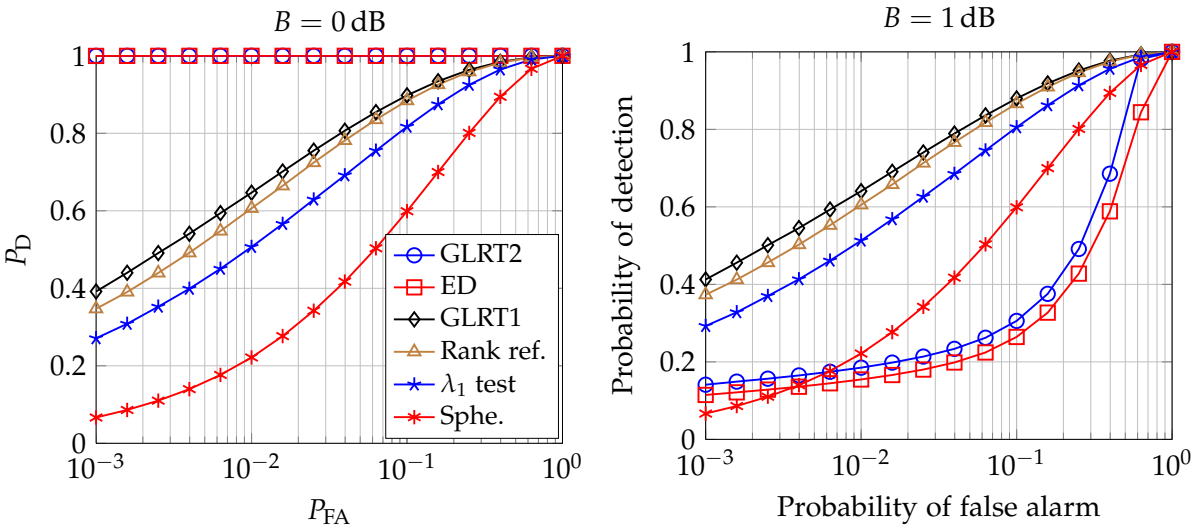

Figure 3.2: ROC curve for different noise uncertainty parameters B.

Figure 3.1 plots the ROC curves, i.e., the detection probability P_D vs. the false alarm rate P_FA for different noise uncertainty parameters B. We assume there are $r = 2$ independent transmit streams sent by the PU. This satisfies the condition $r < M - \sqrt{M}$. The number of sensing samples is $N = 200$. On the one hand, concerning the subplot on the left hand side, the noise variances are perfect known, i.e., $B = 0\,\mathrm{dB}$. It is shown that the GLRT2 method outperforms other methods due to the exploitation of a priori information of noise variances and rank. Without the knowledge of the noise variances, the ED outperforms the other methods, but it is sensitive to the noise uncertainty as shown later. For the other four sensing methods that the noise variances have been estimated, the proposed GLRT1 and the *Rank ref.* algorithm show the best performance since they use the exact rank information to ex-

tract the useful signal subspace from the received signal. Besides, GLRT1, the *Rank ref.* algorithm, and λ_1 test outperform the sphericity test due to the consideration of the unbalanced noises of multiple SUs. On the other hand, the subplot on the right hand side depicts the ROC curves for the noise uncertainty case with $B = 1$ dB. Compared to the left hand side figure, it is observed that the GLRT1, the *Rank ref.* algorithm, the λ_1 test, and the sphericity test are insensitive to such effect, since they estimate the noise variances before applying the GLRT principle. GLRT1 and the *Rank ref.* algorithm outperform other methods because it additionally considers the exact rank information of the signal subspace. Moreover, the proposed GLRT1 and the *Rank ref.* algorithm have the comparably same performance.

Next, we compared the ROC curves of different sensing algorithms in Figure 3.2 under the condition that $r \geq M - \sqrt{M}$. Here we allow for the PUs to transmit $r = 8$ independent streams. The number of sensing samples is $N = 600$. Similar observations are concluded as in Figure 3.1 except that the *Rank ref.* algorithm is suffered from a slight performance loss compared to GLRT1 algorithm. The reason is explained as follow. Under the condition that $r \geq M - \sqrt{M}$, the work [108] stated that the primary signal space has no further special structure besides positive definite Hermitian property; whereas we additionally use the rank information to construct the primary signal space under such circumstances. Thus, the MLE of the unknown parameters here is more accurate, yielding better ROC performance of the GLRT-based methods.

3.3 Spectrum Sensing Under Fading Channels

Cooperative spectrum sensing exploit the multiuser diversity to improve the sensing sensitivity and reliability. However, the performance gain over the local sensing is subject to the fading effect of both the sensing and reporting channels [77]. In order to address this issue, the SUs need to jointly make the sensing decision while exploring the known CSI of the network.

In this section, we study the soft-decision fusion rules for the cooperative sensing. The performance here can be considered as the upper bound achieved by the quantization-based fusion. Both cases, where the sensing and the reporting channels are assumed to be either slow or fast fading channels, are considered. For slow fading channels, the CSI refers to the channel information of block fading channels, while for fast fading channels, the CSI refers to the statistical CSI parameters. We focus on the study in a practical scenario where only the CSI and noise variances of the reporting channels are available. This assumption is justified since the SUs may transmit pilots to the FC for the estimation of the CSI of the reporting channels, while the information of the sensing channels are difficult to obtain due to the lack of cooperation between the PUs and the SUs.

Figure 3.3 depicts a cooperative CR network composed of K SUs and a secondary FC. The cooperative sensing includes two parts: sensing and reporting. In the sensing procedure, each SU performs spectrum sensing based on the observation over N sam-

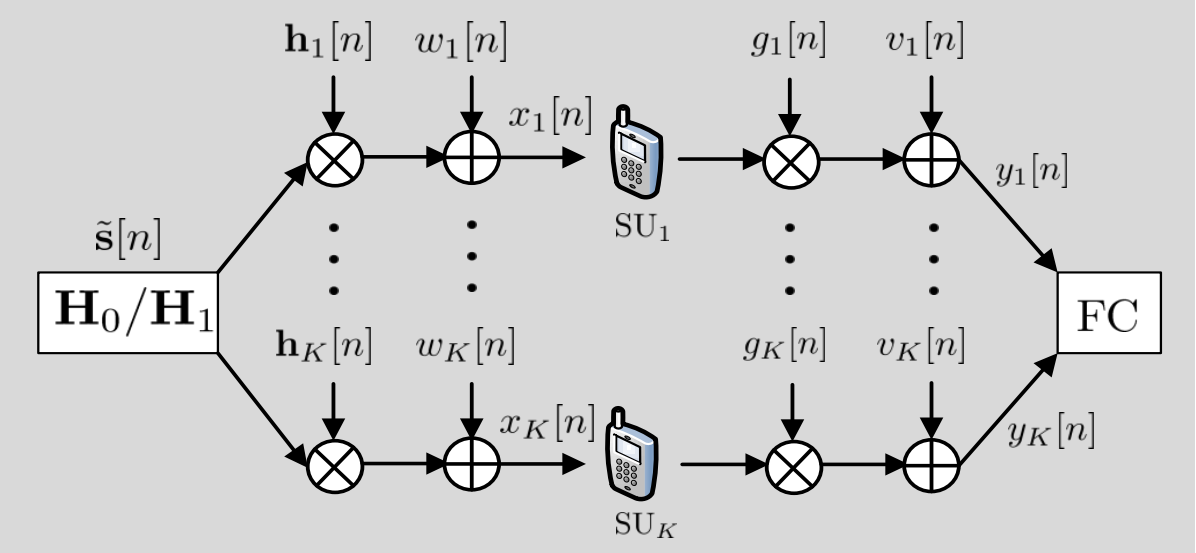

Figure 3.3: Cooperative sensing in the secondary network with K SUs and FC at time instant n.

pling times. Under $\mathcal{H}_1$, the primary signal is assumed to be a zero-mean temporally and spatially white signal with unit power for each transmit antenna. The primary transmit signal at the nth time instant is $\tilde{\mathbf{s}}[n] \in \mathbb{C}^{N_t \times 1}$ where N_t is the number of transmit antennas at the PU. Assuming each SU has a single receive antenna and denoting $\mathbf{h}_k[n] \in \mathbb{C}^{1 \times N_t}$ as the stationary Rayleigh fading channel from the PU to the kth SU at the nth instant, we obtain the matrix $\mathbf{H}[n] = \left[\mathbf{h}_1[n]^T, \ldots, \mathbf{h}_K[n]^T\right]^T$. The signal part received by the SUs at the nth instant is

$$\mathbf{s}[n] = \mathbf{H}[n]\tilde{\mathbf{s}}[n] \tag{3.43}$$

and $\mathbf{\Sigma}_S$ is the covariance matrix of $\mathbf{s}[n]$. Assuming that $x_k[n]$ is the nth received signal sample at the kth SU, the received signals at K SUs are stacked into a vector $\mathbf{x}[n] = [x_1[n], \ldots, x_K[n]]^T$ as

$$\mathbf{x}[n] = \begin{cases} \mathbf{w}[n], & \mathcal{H}_0 \\ \mathbf{s}[n] + \mathbf{w}[n], & \mathcal{H}_1 \end{cases}, \quad \forall\, n = 1, \ldots, N \tag{3.44}$$

where the temporally and spatially white noise is $\mathbf{w}[n] = [w_1[n], \ldots, w_K[n]]^T$ with $n_k[n], \forall\, k = 1, \ldots, K$ being the noise sample for the kth SU at the nth time instant. We assume $\mathbf{w}[n] \sim \mathcal{CN}(0, \sigma_w^2 \mathbf{I}_K)$. The signal samples $\{\mathbf{s}[n]\}_{n=1}^N$ and the noise samples $\{\mathbf{w}[n]\}_{n=1}^N$ are mutually independent.

In the reporting procedure, the FC receives the signals relayed by each SU through the reporting channels in an orthogonal manner to exploit the spatial diversity [106]

$$\mathbf{y}[n] = \mathbf{G}[n]\mathbf{x}[n] + \mathbf{v}[n] \tag{3.45}$$

where $\mathbf{G} = \text{diag}\{\mathbf{g}[n]\}$ with

$$\mathbf{g}[n] = [g_1[n], \ldots, g_K[n]]^T$$

and $g_k[n]$ representing the fading coefficient from the kth SU to the FC. We remark that $g_k[n]$ is not restricted to Rayleigh fading channels, e.g., a more general Rician fading channel model can be applied if there is a line of sight between the SU and the FC. $\mathbf{v}[n] = [v_1[n], \ldots, v_K[n]]^T$ is the vector containing the temporally and spatially white noise samples of the reporting channel and $\mathbf{v}[n] \sim \mathcal{CN}(0, \sigma_v^2 \mathbf{I}_K)$.

Two kinds of fading models are applied to sensing and reporting channels. We also assume different types of available CSI for two channel models, respectively.

- **Fast fading channels**: The instantaneous CSI varies for different time instants and is hard to obtain. Thus, only statistical CSI is assumed to be available. We use $\mathbf{\Sigma}_H = \mathbb{E}\{\mathbf{H}[n]\mathbf{H}[n]^H\}$, $\mathbf{g}[n] \sim \mathcal{CN}(\bar{\mathbf{g}}, \mathbf{\Sigma}_g)$. Consequently, we obtain

$$\mathbf{\Sigma}_S = \mathbf{\Sigma}_H, \quad \mathbf{R}_g = \mathbf{\Sigma}_g + \bar{\mathbf{g}}\bar{\mathbf{g}}^H \tag{3.46}$$

  where $\mathbf{R}_g$ is the correlation matrix of $\mathbf{g}[n]$.

- **Slow fading channels**: The CSI $\{\mathbf{H}[n]\}$ and $\{\mathbf{g}[n]\}$ are the block fading values that are fixed for $\forall\, n = 1, \ldots, N$. Therefore, the time argument n in $\mathbf{H}[n]$ and $\mathbf{g}[n]$ is dropped for simplicity. We obtain

$$\mathbf{\Sigma}_S = \mathbf{H}\mathbf{H}^H, \quad \mathbf{R}_g = \mathbf{g}\mathbf{g}^H. \tag{3.47}$$

We aim at designing the sensing rule at the FC to combine $\mathbf{y}[n]$, $\forall\, n = 1, \ldots, N$. According to the Neyman-Pearson optimality criterion, for a binary hypothesis testing problem, the optimal test statistic to maximize the P_{D} given a fixed P_{FA} is the following log LRT (LLRT) scheme [66]:

$$\Lambda_{\text{LLRT}}(\tilde{\mathbf{y}}) = \ln \frac{p(\tilde{\mathbf{y}}|\mathcal{H}_1)}{p(\tilde{\mathbf{y}}|\mathcal{H}_0)} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \gamma \tag{3.48}$$

where γ is the threshold and

$$\tilde{\mathbf{y}} = \left[\mathbf{y}^T(1), \ldots, \mathbf{y}^T(N)\right]^T.$$

The conditional PDF $p(\tilde{\mathbf{y}}|\mathcal{H}_i)$, $i = 0, 1$ is difficult to calculate since each $\mathbf{y}[n]$ contains, e.g., a multiplication of two Gaussian distributions [53]. Instead, we approximate $p(\tilde{\mathbf{y}}|\mathcal{H}_i)$ by using a Gaussian distribution $p_G(\tilde{\mathbf{y}}|\mathcal{H}_i)$. Since the observations $\{\mathbf{y}[n]\}_{n=1}^N$ are independent with the Gaussian assumption, we have

$$p_G(\tilde{\mathbf{y}}|\mathcal{H}_i) = \prod_{n=1}^N p_G(\mathbf{y}[n]|\mathcal{H}_i).$$

Dropping the time argument n for $\mathbf{y}[n]$, we obtain

$$p_G(\mathbf{y}|\mathcal{H}_i) \sim \mathcal{CN}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$$

where $\boldsymbol{\mu}_i$ and $\boldsymbol{\Sigma}_i$ are

$$\boldsymbol{\mu}_i = \mathbb{E}\{\mathbf{y}|\mathcal{H}_i\} \tag{3.49}$$
$$\boldsymbol{\Sigma}_i = \mathbb{E}\{\mathbf{y}\mathbf{y}^H|\mathcal{H}_i\} - \boldsymbol{\mu}_i\boldsymbol{\mu}_i^H. \tag{3.50}$$

The closed-form expressions of (3.49) and (3.50) are calculated in Appendix A.1. The results are

$$[\boldsymbol{\mu}_i]_k = \bar{g}_k\bar{x}_{k,i} \tag{3.51}$$
$$[\boldsymbol{\Sigma}_i]_{k,k'} = \sigma_v^2\delta_{k,k'} + \mathbb{E}(x_k x_{k'}^*)\mathbb{E}(g_k g_{k'}^*) - [\boldsymbol{\mu}_i]_k[\boldsymbol{\mu}_i]_{k'}^* \tag{3.52}$$

with

$$\bar{g}_k = \mathbb{E}(g_k)$$
$$\bar{x}_{k,i} = \mathbb{E}(x_k|\mathcal{H}_i).$$

According to (3.44), the received signal $\mathbf{x}$ can be approximated by the following Gaussian distribution

$$\mathbf{x} \sim \begin{cases} \mathcal{CN}(\mathbf{0}, \sigma_w^2\mathbf{I}_K), & \mathcal{H}_0 \\ \mathcal{CN}(\mathbf{0}, \boldsymbol{\Sigma}_S + \sigma_w^2\mathbf{I}_K), & \mathcal{H}_1. \end{cases} \tag{3.53}$$

Integrating (3.53) into (3.51) and (3.52), the mean and covariance matrices of $\mathbf{y}$ are obtained as

$$\boldsymbol{\mu}_0 = \mathbf{0}, \quad \boldsymbol{\Sigma}_0 = \sigma_w^2\mathbf{I}_K \odot \mathbf{R}_g + \sigma_v^2\mathbf{I}_K \tag{3.54}$$
$$\boldsymbol{\mu}_1 = \mathbf{0}, \quad \boldsymbol{\Sigma}_1 = (\boldsymbol{\Sigma}_S + \sigma_w^2\mathbf{I}_K) \odot \mathbf{R}_g + \sigma_v^2\mathbf{I}_K \tag{3.55}$$

where $\boldsymbol{\Sigma}_\mathbf{S}$ and $\mathbf{R}_g$ are given in (3.46) and (3.47) for two kinds of fading channels, respectively.

3.3.1 Approximated Likelihood Ratio Test

Due to the lack of the analytical form of the exact PDF $p(\tilde{\mathbf{y}}|\mathcal{H}_i)$, $i = 1, 2$, we replace it with the Gaussian approximation $p_G(\tilde{\mathbf{y}}|\mathcal{H}_i)$. The resulting near-optimal method is named approximated likelihood ratio test (ALRT) method under the assumptions of the known CSI and noise variances of both sensing and reporting channels.

In particular, substituting $p_G(\tilde{\mathbf{y}}|\mathcal{H}_i)$ into (3.48) and integrating the constant items into the threshold design, the test statistic of the ALRT detector is given as

$$\Lambda_{\text{ALRT}}(\tilde{\mathbf{y}}) = \sum_{n=1}^{N} \mathbf{y}^H[n]\left(\boldsymbol{\Sigma}_0^{-1} - \boldsymbol{\Sigma}_1^{-1}\right)\mathbf{y}[n]. \tag{3.56}$$

3.3.2 Energy Detection

If the FC sums up the energy of the signals relayed by each SU, the detection scheme becomes the widely-used ED

$$\Lambda_{\mathrm{ED}}(\tilde{\mathbf{y}}) = \sum_{n=1}^{N} \mathbf{y}^H[n]\mathbf{y}[n]. \tag{3.57}$$

The test statistic Λ_{ED} can be considered as the special case of Λ_{ALRT} when $\left(\boldsymbol{\Sigma}_0^{-1} - \boldsymbol{\Sigma}_1^{-1}\right)$ is equal to an identity matrix. The ED only requires knowledge of the primary signal such as the center frequency and bandwidth, but neither *a priori* information of the signal structure nor of the CSI.

3.3.3 GLRT-Based Detection

Given the approximated signal model, we apply the GLRT principle (3.3) and yield

$$\Lambda_{\mathrm{GLRT}}(\tilde{\mathbf{y}}) = \ln \frac{p_G(\tilde{\mathbf{y}}|\mathcal{H}_1, \hat{\theta}_1)}{p_G(\tilde{\mathbf{y}}|\mathcal{H}_0, \hat{\theta}_0)} \tag{3.58}$$

with

$$\hat{\theta}_0 = \arg\max_{\theta_0} p_G(\tilde{\mathbf{y}}|\mathcal{H}_0, \theta_0) \tag{3.59}$$

$$\hat{\theta}_1 = \arg\max_{\theta_1} p_G(\tilde{\mathbf{y}}|\mathcal{H}_1, \theta_1). \tag{3.60}$$

Note that the exact distribution $p(\tilde{\mathbf{y}}|\mathcal{H}_i)$ with $i = 0, 1$ is replaced with the approximated Gaussian distribution $p_G(\tilde{\mathbf{y}}|\mathcal{H}_i)$ for the sake of computational tractability.

3.3.3.1 Unknown Noise Variances

In this case, both noise variance σ_w^2 of the sensing channels and $\boldsymbol{\Sigma}_S$ need to be estimated under both hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$.

First, the LLF under $\mathcal{H}_0$ with the unknown σ_w^2 is written as

$$\ln p_G(\tilde{\mathbf{y}}|\mathcal{H}_0) = -KN\ln\pi - N\ln\det\boldsymbol{\Sigma}_0 - N\mathrm{tr}\left\{\boldsymbol{\Sigma}_0^{-1}\mathbf{R}_{\bar{y}}\right\}. \tag{3.61}$$

where

$$\mathbf{R}_{\bar{y}} = \frac{1}{N}\sum_{n=1}^{N} \mathbf{y}[n]\mathbf{y}^H[n] \overset{(a)}{=} \mathbf{U}_{\bar{y}}\boldsymbol{\Lambda}_{\bar{y}}\mathbf{U}_{\bar{y}}^H. \tag{3.62}$$

The step (a) in (3.62) indicates the eigenvalue decomposition of $\mathbf{R}_{\bar{y}}$. The term

$$\boldsymbol{\Lambda}_{\bar{y}} = \mathrm{diag}\left\{\lambda_{1,\bar{y}}, \ldots, \lambda_{K,\bar{y}}\right\}$$

contains all the eigenvalues of $\mathbf{R}_{\bar{y}}$ as diagonal entries in descending order. After some mathematical manipulation, the MLE of σ_w^2 is derived as

$$\widehat{\sigma_w^2}|\mathcal{H}_0 = \frac{1}{K}\text{tr}\{\mathbf{C}\} \tag{3.63}$$

with

$$[\mathbf{C}]_{i,j} = \frac{\left[\mathbf{U}_{\bar{y}}\text{diag}\left\{\left(\lambda_{1,\bar{y}} - \sigma_v^2\right)^+, \ldots, \left(\lambda_{K,\bar{y}} - \sigma_v^2\right)^+\right\}\mathbf{U}_{\bar{y}}^H\right]_{i,j}}{\left[\mathbf{R}_g\right]_{i,j}}$$

Similarly, the LLF under $\mathcal{H}_1$ with the unknown $\mathbf{\Sigma}_S$ and σ_w^2 is

$$\ln p_G(\tilde{\mathbf{y}}|\mathcal{H}_1) = -KN\ln\pi - N\ln\det\mathbf{\Sigma}_1 - N\text{tr}\left\{\mathbf{\Sigma}_1^{-1}\mathbf{R}_{\bar{y}}\right\}. \tag{3.64}$$

The MLE of the unknown parameters under $\mathcal{H}_1$ is

$$\widehat{\mathbf{\Sigma}_S + \sigma_w^2}\mathbf{I}_K|\mathcal{H}_1 = \mathbf{C}. \tag{3.65}$$

Integrating (3.63) and (3.65) into (3.61) and (3.64), the GLRT test statistic is given as

$$\Lambda_{\text{GLRT3}}(\tilde{\mathbf{y}}) = -N\ln\prod_{k=1}^{m_3}\frac{\lambda_{k,\bar{y}}}{\sigma_v^2} - Nm_3 - N\sum_{k=m_3+1}^{K}\frac{\lambda_{k,\bar{y}}}{\sigma_v^2} \tag{3.66}$$

$$+ N\ln\prod_{k=1}^{K}\lambda_{k,e} + N\text{tr}\left\{\mathbf{E}^{-1}\mathbf{R}_{\bar{y}}\right\} \tag{3.67}$$

where

$$\mathbf{E} = \left(\widehat{\sigma_w^2}|\mathcal{H}_0\right)\mathbf{I}_K \odot \mathbf{R}_g + \sigma_v^2\mathbf{I}_K$$

and $\lambda_{k,e}$, $\forall\, k = 1, \ldots, K$, are the eigenvalues of $\mathbf{E}$. The value m_3 refers to the largest m_3 such that

$$\lambda_{m_3,\bar{y}} \geq \sigma_v^2.$$

3.3.3.2 Known Noise Variances

If the covariance matrix $\mathbf{\Sigma}_S$ is unknown, i.e., the power of the primary signal and the CSI of the sensing channel are unknown, they can be estimated by exploiting the structure of the sample covariance matrix of the received signals. Specifically, the LLF under $\mathcal{H}_0$ is

$$\ln p_G(\tilde{\mathbf{y}}|\mathcal{H}_0) = -KN\ln\pi - N\ln\det\mathbf{\Sigma}_0 - \sum_{n=1}^{N}\mathbf{y}^H[n]\mathbf{\Sigma}_0^{-1}\mathbf{y}[n] \tag{3.68}$$

where all parameters in (3.68) are known, i.e., there is no parameter to be estimated in (3.58). By using a decomposition of the form

$$\mathbf{\Sigma}_0 = \mathbf{L}_1^H \mathbf{L}_1$$

(3.68) is reformulated as

$$\ln p_G(\tilde{\mathbf{y}}|\mathcal{H}_0) = -KN \ln \pi - N \ln \det \mathbf{L}_1^H \mathbf{L}_1 - N \mathrm{tr}\left\{\mathbf{R}_y\right\} \tag{3.69}$$

where

$$\mathbf{R}_y = \mathbf{L}_1^{-H} \left(\frac{1}{N} \sum_{n=1}^{N} \mathbf{y}[n] \mathbf{y}^H[n] \right) \mathbf{L}_1^{-1}.$$

Similarly, the LLF under $\mathcal{H}_1$ is reformulated as follows due to the temporal whiteness of $\tilde{\mathbf{s}}[n]$, $\forall\, n = 1, \ldots, N$

$$\ln p_G(\tilde{\mathbf{y}}|\mathcal{H}_1) = -KN \ln \pi + N \ln \det \left(\mathbf{L}_1^{-1} \mathbf{B} \mathbf{L}_1^{-H} \right) - N \mathrm{tr}\left\{\mathbf{R}_y \mathbf{B}\right\}. \tag{3.70}$$

where

$$\mathbf{B} = \left(\mathbf{L}_1^{-H} \left(\mathbf{\Sigma}_S \odot \mathbf{R}_g \right) \mathbf{L}_1^{-1} + \mathbf{I}_K \right)^{-1}. \tag{3.71}$$

According to (3.71), the unknown parameter $\mathbf{\Sigma}_S$ is only included in $\mathbf{B}$, i.e., the unknown parameter θ_1 in (3.58) is $\mathbf{B}$. The MLE of $\mathbf{B}$ is obtained by solving the following constrained optimization problem

$$\begin{aligned} \max_{\mathbf{B}} \quad & \ln \det \mathbf{B} - \mathrm{tr}\left\{\mathbf{R}_y \mathbf{B}\right\} \\ \text{s.t.} \quad & \mathbf{0} \preceq \mathbf{B} \preceq \mathbf{I}_K. \end{aligned}$$

Similar to [148, Appendix], we obtain the MLE of $\mathbf{B}$

$$\hat{\mathbf{B}} = \mathbf{U}_y \mathrm{diag}\left\{ \min\left(\lambda_{1,y}^{-1}, 1\right), \ldots, \min\left(\lambda_{K,y}^{-1}, 1\right) \right\} \mathbf{U}_y^H \tag{3.72}$$

where $\mathbf{U}_y$ contains the eigenvectors of $\mathbf{R}_y$ and $\lambda_{k,y}$, $\forall\, k = 1, \ldots, K$ are the corresponding K eigenvalues listed in descending order, i.e., $\lambda_{1,y} \geq \lambda_{2,y} \geq \ldots \geq \lambda_{K,y}$.

Taking the MLE result (3.72) into (3.70) and subtracting (3.69), it results in the test statistic of GLRT4

$$\mathbf{\Lambda}_{GLRT4}(\tilde{\mathbf{y}}) = N \ln \det \hat{\mathbf{B}} - N \mathrm{tr}\left\{\mathbf{R}_y \hat{\mathbf{B}}\right\} + N \mathrm{tr}\left\{\mathbf{R}_y\right\}. \tag{3.73}$$

Using (3.72), we reformulate the two terms containing $\hat{\mathbf{B}}$ as

$$\ln \det \hat{\mathbf{B}} = -\ln \prod_{k=1}^{m_4} \lambda_{k,y} \tag{3.74}$$

$$\mathrm{tr}\left\{\mathbf{R}_y\hat{\mathbf{B}}\right\} = m_4 + \sum_{k=m_4+1}^{K} \lambda_{k,y} \tag{3.75}$$

where m_4 refers to the largest m_4 such that $\lambda_{m_4,y} \geq 1$. Integrating (3.74) and (3.75) into (3.73) results in the final test statistic of GLRT4

$$\mathbf{\Lambda}_{\mathrm{GLRT4}}(\tilde{\mathbf{y}}) = Nm_4 \left(\frac{1}{m_4} \sum_{k=1}^{m_4} \lambda_{k,y} - \ln \left(\prod_{k=1}^{m_4} \lambda_{k,y} \right)^{\frac{1}{m_4}} - 1 \right). \tag{3.76}$$

Remark: The performance of the GLRT-based methods depends on the distinguishable statistical properties of the estimated parameters under both hypotheses. Specifically, the following two properties are exploited to differentiate binary hypotheses : under the hypothesis $\mathcal{H}_1$, the received signals from multiple SUs at the FC are correlated and have unbalanced power levels due to different properties of the sensing and the reporting channels; on the contrary, under the hypothesis $\mathcal{H}_0$, the received noises are uncorrelated and their variances are same. However, in fast fading scenario, if the reporting channel has low spatial correlation, i.e., $\mathbf{R}_g$ is a scaled identity matrix, the correlation effect diminishes at the FC according to (3.54) and (3.55). This results in a performance degradation of the GLRT-based methods, since only the property of unbalanced power levels is used to identify the presence and absence of the PU signal. Nevertheless, the performance can be improved by other approaches, e.g., prolonging the sensing length.

3.3.4 Performance Evaluation

To numerically evaluate the performance, we consider a CR network with $K = 4$ SUs detecting one PU with a single transmit antenna. The primary signals are quadrature phase-shift keying modulated signals with unit variance. Each channel $\mathbf{h}_k[n]$ or $\mathbf{g}_k[n]$ is assumed to be a Rayleigh fading channel. The SNRs of the sensing and the reporting channels are specified in each figure. The noise variance of the reporting links is $\sigma_v^2 = 0.8$. The noise uncertainty effect of the sensing links is considered, i.e., the estimated noise variance is $\sigma_w^2 = 1.5$ and the true noise variance is $\tilde{\sigma}_w^2 = \sigma_w^2/\alpha$, where $10\log_{10}\alpha$ is uniformly distributed in $[-B, B]$ with noise uncertainty factor B in dB [113]. For example, $B = 0\,\mathrm{dB}$ indicates that the noise variance is precisely known.

We compare the following four sensing algorithms.

- ALRT: The test statistic is given in (3.56).
- ED: The test statistic is given in (3.57).
- GLRT3: The test statistic is given in (3.67).
- GLRT4: The test statistic is given in (3.76).

Figure 3.4 plots the ROC curves for slow fading channels. The observation length is $M = 1000$. Two cases of noise uncertainty are considered: $B = 0$ dB and $B = 1$ dB.

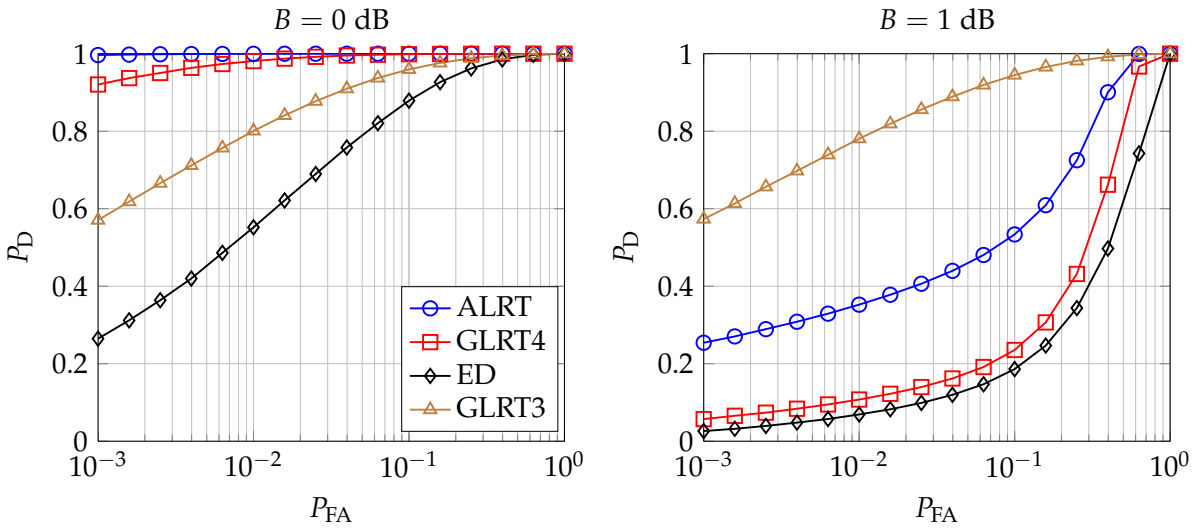

Figure 3.4: ROC curve with noise uncertainty for slow fading channels. The SNRs of the sensing links are set to [-18, -11, -16, -10] dB and the SNRs of the reporting links are [8, 10, 12, 6] dB, M=1000.

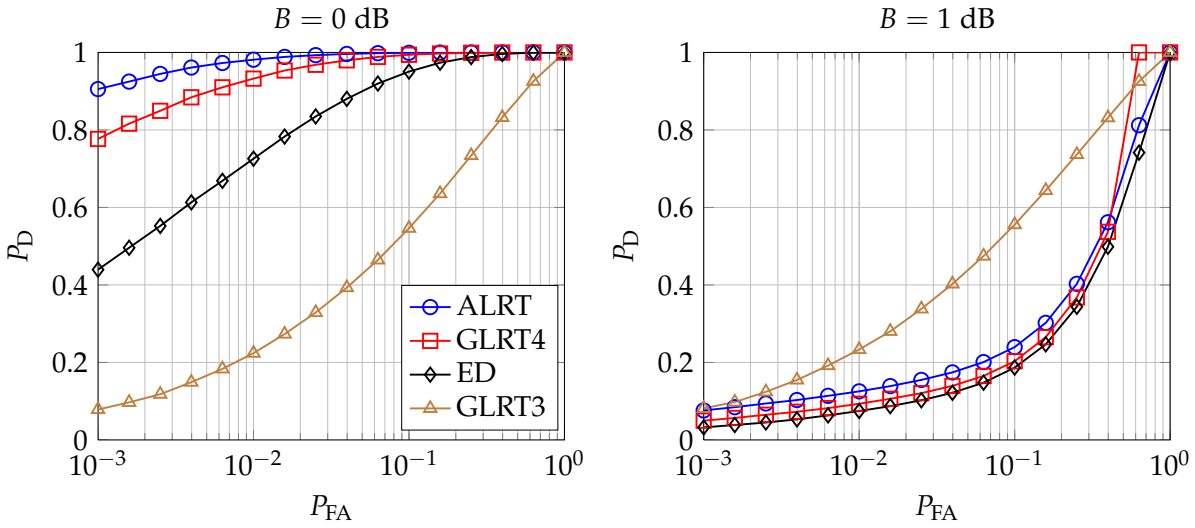

Figure 3.5: ROC curve with noise uncertainty for fast fading channels. The SNRs of the sensing links are set to [-18, -11, -16, -10] dB and the SNRs of the reporting links are [8, 10, 12, 6] dB, M=3000.

In the legend, the value of B is denoted in parentheses. We observe that if the noise variance σ_w^2 is perfectly known, the ALRT method performs the best due to complete a priori knowledge. Without the knowledge of $\mathbf{\Sigma}_S$, GLRT4 uses the hidden knowledge in the received signals and outperforms the ED. If a noise uncertainty exists, GLRT3 always performs the best in the region of interest since the noise variance is estimated. As expected, GLRT3 is shown to be insensitive to the uncertainty, while

the performance of the other three methods degrades severely. Figure 3.5 depicts the ROC curves for fast fading channels. The corresponding correlation matrix $\mathbf{\Sigma}_H$ or $\mathbf{\Sigma}_g$ is a Toeplitz matrix $[\mathbf{\Sigma}_H]_{i,j} = 0.1^{|i-j|}$ and $[\mathbf{\Sigma}_g]_{i,j} = 0.1^{|i-j|}$, respectively. The observation length is $M = 3000$. Similar observations are obtained as in Figure 3.4.

3.4 Summary

In CR networks, cooperative spectrum sensing is more favorable than local spectrum sensing due to its improved sensitivity and the effectiveness in avoiding the hidden terminal problem. However, the performance of cooperative sensing is subject to several realistic restrictions. For example, it is difficult for the SUs to acquire perfect knowledge of the primary signal space and the CSI of the entire network. Furthermore, the fading of the sensing and reporting channels can deteriorate the sensing performance. Hence, we have investigated effective spectrum sensing methods aiming at tackling the aforementioned challenges.

The spectrum sensing problems with limited a priori knowledge can be modeled as binary hypothesis testing problems with unknown parameters. In this chapter, we have applied the GLRT principle to two realistic sensing examples. Firstly, assuming an unknown structure of the primary signal space at the SUs, we exploited the rank information to extract its structure. The GLRT-based sensing methods were then designed corresponding to two scenarios, i.e., unknown or known noise variances. The resulting algorithms also generalize some special cases in the literature. Secondly, we considered the influence of the fading sensing and reporting channels on cooperative sensing. The SUs are assumed to only have partial CSI of the channels but without knowing the structure of the primary signal space. We derived the sensing algorithms under two scenarios, i.e., the noise variances of the sensing channels are unknown or known. Overall, the proposed GLRT-based algorithms effectively utilize the available information at the SUs in countering the performance degradation caused by environment uncertainties, e.g., fading effects and noise uncertainty.

Chapter 4

Power Allocation with Partial Primary CSI

The underlay paradigm allows for the simultaneous transmission of PUs and SUs. Due to the hierarchical usage of the resources, the main goal is to design the secondary transceiver strategies in order to constrain the performance degradation caused to the PUs below some tolerable limit.

Power allocation is an effective way for interference mitigation. Several works [8, 33, 37, 64, 120] have considered the design of power control strategies using the conventional metric to evaluate the performance degradation of the PU, i.e., the IT constraint denoting the received interference power at the PR. Alternatively, limiting the primary capacity loss is the ultimate goal in regulating the secondary transmission. To this end, the authors in [65, 149] demonstrated that a gain on the secondary rate is achieved by further exploiting the primary CSI instead of considering solely the IT constraint during the system design.

In practice, the SUs can only acquire partial CSI of the entire network, especially regarding the CSI related to the PUs. Two kinds of partial CSI models were considered in the literature: outdated CSI [73, 89, 94, 110, 121] and statistical CSI [12, 28, 80, 116]. On the one hand, outdated CSI is obtained on a real-time basis subject to the delay caused in channel feedback or estimation. The works [73, 89] found that the low correlation between the outdated and the actual CSI yields a large performance loss of the secondary transmission. On the other hand, statistical CSI changes on a time scale that is much larger than the channel coherence time. Given such CSI, the performance loss of the PUs can be limited in an average or a probability-constrained manner.

Given the power control strategies, it is desirable to obtain insights on the achievable performance of the SUs at the expense of the performance degradation to the PUs. Such performance analysis can be used for a performance assessment and to provide guidelines in selecting system parameters for power adaptation strategies. However, the optimal power strategy is usually not given in closed form. Thus, it brings in mathematical challenges in the performance analysis. An alternative way is to develop a near-optimal strategy. If the corresponding analytical performance is obtained, it can be a good approximation of the optimal achievable performance.

In this chapter, power allocation strategies for the secondary transmission are investigated subject to different QoS constraints on the primary link under Rayleigh fading channels assuming only partial CSI related to the PR is available at the ST. After introducing the system model in Section 4.1, we study the power allocation in Section 4.2 subject to an average IT constraint and a transmit power constraint on the ST. Motivated by the benefit of exploiting the primary CSI [149], the power allocation

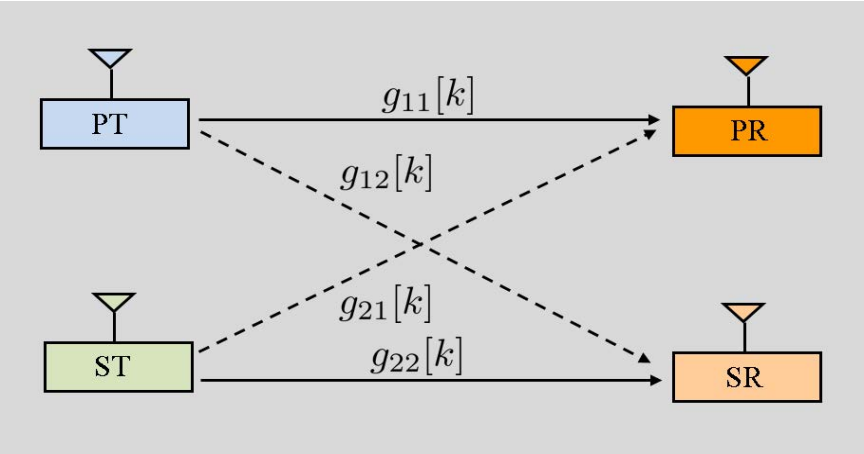

Figure 4.1: Spectrum sharing system with a primary and a secondary link.

strategies are designed in Section 4.3 using the statistical CSI of the primary link subject to an outage probability constraint on the PU instead of an IT constraint. We aim at not only deriving the optimal power allocation strategies but also designing the low-complexity near-optimal strategies with the corresponding performance analysis.

The results presented in this chapter have been published in part by the author in [40,41,47–49][1].

4.1 System Model

Figure 4.1 depicts the considered cross-interfering spectrum sharing system in which a secondary link coexists with a primary link. We assume that a single antenna is equipped at the ST, the SR, the PT, and the PR, respectively. The transmit symbols of the PT and the ST at the kth time instant are given by $x_1[k] \sim \mathcal{CN}(0,1)$ and $x_2[k] \sim \mathcal{CN}(0,1)$, respectively. We assume that the PT uses non-adaptive power transmission with the power denoted by P_1, while the ST optimizes the power $P_2[k]$ according to the CSI at the kth time instant. We remark that similar to [37,69], here the term "power" refers to the instantaneous power averaged over the transmit symbols.

All channels are assumed to be stationary, ergodic, and mutually independent Rayleigh flat fading channels in a slow-fading scenario. Instantaneous CSI of the PT-PR link, the PT-SR link, the ST-PR link, and the ST-SR link is given by $h_{11}[k]$, $h_{12}[k]$, $h_{21}[k]$, and $h_{22}[k]$, respectively. The noise of the primary and the secondary link is given by $n_p[k] \sim \mathcal{CN}(0,\sigma_P^2)$ and $n_s[k] \sim \mathcal{CN}(0,\sigma_S^2)$, respectively, with positive and

[1] In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of RWTH Aachen University's products or services. Internal or personal use of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution, please go to `http://www.ieee.org/publications_standards/publications/rights/rights_link.html` to learn how to obtain a License from RightsLink.

finite σ_P^2 and σ_S^2. Consequently, the received signals at the PR and the SR at the kth time instant are denoted by $y_1[k]$ and $y_2[k]$, respectively:

$$y_1[k] = \sqrt{P_1}h_{11}[k]x_1[k] + \sqrt{P_2[k]}h_{21}[k]x_2[k] + n_p[k] \tag{4.1}$$

$$y_2[k] = \sqrt{P_2[k]}h_{22}[k]x_2[k] + \sqrt{P_1}h_{12}[k]x_1[k] + n_s[k]. \tag{4.2}$$

The channel power gain is denoted by

$$g_{ij}[k] = \left|h_{ij}[k]\right|^2, \;\; i, j = 1, 2 \tag{4.3}$$

which is exponentially distributed with the PDF

$$f_{g_{ij}}(x) = \begin{cases} l_{ij}e^{-l_{ij}x}, & x \geq 0 \\ 0, & x < 0 \end{cases} \quad i, j = 1, 2 \tag{4.4}$$

with l_{ij} indicating the rate parameter. In practice, the instantaneous CSI $g_{11}[k]$ and $g_{21}[k]$ is difficult to obtain at the ST due to limited cooperation between the SU and the PU. However, we assume statistical parameters l_{11} and l_{21} to be available at the ST due to the exploitation of side information [61]. For the PT-SR link, either statistical CSI parameter l_{12} or the the instantaneous CSI $g_{12}[k]$ is available at the ST, depending on the scenario whether the ST can obtain statistical parameters by the location information or obtain the instantaneous CSI via estimation or feedback. Additionally, we assume both the ST and the SR know perfect instantaneous CSI of the secondary link $g_{22}[k]$ and the SR additionally knows the instantaneous CSI of the PT-SR link $g_{12}[k]$. For the remainder of the chapter, we omit the time argument k for simplicity.

We consider two kinds of QoS constraints to restrict the performance degradation of the primary transmission: average IT constraint and outage probability constraints. Besides, the peak power of the ST is limited. The power control strategies aiming at maximizing the achievable rate of the secondary link subject to aforementioned constraints are investigated in the remainder of this chapter.

4.2 Power Control with Interference Power Constraint

4.2.1 Known CSI of the PT-SR Link

Recalling the system model in Figure 4.1, we assume the ST and the SR know instantaneous CSI g_{12} and g_{22}, but only statistical CSI parameter l_{21}. This assumption is justified due to the fact that such instantaneous CSI can be possibly obtained in two ways: First, the SR can estimate g_{12} and g_{22} by exploiting the pilot symbols from the ST and the PT and then feedback it to the ST; second, the ST can also estimate such CSI by exploiting the reciprocity in the TDD transmission.

Two power constraints are considered. First, the average interference power constraint P_I is imposed concerning the protection of the PU. Second, the peak power con-

straint P_S is set for the ST. We assume Gaussian input signallings and single-user detection at the SR in which the interfering signal from the PT is taken as noise [17,126], The optimal power strategy subject to both power constraints is obtained by solving the following optimization problem

$$\begin{aligned} \max_{P_2(g_{12},\, g_{22})} \quad & \mathbb{E}_{g_{12},\, g_{22}} \ln\left(1+\frac{P_2(g_{12},\, g_{22})g_{22}}{P_1 g_{12}+\sigma_S^2}\right) \\ \text{s.t.} \quad & P_2(g_{12},g_{22}) \le P_S, \quad \forall\, g_{12}\ge 0, g_{22}\ge 0 \\ & P_2(g_{12},g_{22}) \ge 0, \quad \forall\, g_{12}\ge 0, g_{22}\ge 0 \\ & \mathbb{E}_{g_{12},\, g_{22},\, g_{21}}\left\{P_2(g_{12},g_{22})g_{21}\right\} \le P_I \end{aligned} \tag{4.5}$$

where the notation $P_2(g_{12},\, g_{22})$ is used because the power is adapted to the instantaneous knowledge g_{12} and g_{22}. The unit of the objective function is nats/s/Hz.

Since the ST only has statistical CSI of the ST-PR link, the optimal power allocation strategy is independent from each instantaneous channel realization g_{21}. Considering the mutual independence of the CSI of each link, we convert the third constraint of (4.5) to the following inequality

$$\mathbb{E}_{g_{12},\, g_{22}}\left\{P_2(g_{12},g_{22})\right\} \le P_I l_{21}. \tag{4.6}$$

For the sake of simplicity, introducing the variable t as

$$t=\frac{\sigma_S^2+P_1 g_{12}}{g_{22}} \tag{4.7}$$

we represent $P_2(g_{12},g_{22})$ as $P_2(t)$ and reformulate the optimization problem (4.5) as

$$\begin{aligned} \max_{P_2(t)} \quad & R\left(P_2(t)\right) \\ \text{s.t.} \quad & P_2(t) \le P_S, \quad \forall \quad t>0 \\ & P_2(t) \ge 0, \quad \forall \quad t>0 \\ & \mathbb{E}_t\left\{P_2(t)\right\} \le P_I l_{21} \end{aligned} \tag{4.8}$$

where $R\left(\cdot\right)$ is defined as

$$R\left(x(t)\right)=\mathbb{E}_t\left\{\ln\left(1+\frac{x(t)}{t}\right)\right\}. \tag{4.9}$$

Note that since the optimization variable $P_2(t)$ in (4.8) is a function instead of a vector in Euclidean space, the optimization theory in function space [58,85] is applied later to derive the optimal solution. A brief revision of KKT conditions for functional optimization is given in [87].

4.2.1.1 Optimal Power Allocation

The optimization problem (4.8) is a constrained convex problem and the Slater condition is satisfied. Then the KKT conditions are sufficient and necessary conditions, i.e., the KKT solution is the global optimal solution [111]. This argument provides the theoretical basis to design the optimization strategy.

Proposition 4.2.1. *The KKT solution to the optimization problem (4.8) is*

$$P_2^\star(t) = \begin{cases} 0, & t \geq v^\star \\ v^\star - t, & v^\star - P_S < t < v^\star \\ P_S, & 0 \leq t \leq v^\star - P_S. \end{cases} \tag{4.10}$$

The non-negative parameter $v^\star$ is the inverse of an optimal Lagrangian multiplier $\lambda^\star$ related to the constraint (4.6). It is chosen to fulfill the average power constraint (4.6).

Proof. The Lagrangian of (4.8) is written as

$$L = \mathbb{E}_t \left\{ \ln \left(1 + \frac{P_2(t)}{t} \right) \right\} + \lambda \left(\mathbb{E}_t \left\{ P_2(t) \right\} - P_I l_{21} \right) \tag{4.11}$$

where the non-negative Lagrangian multiplier λ corresponds to the constraint (4.6). Herein, no Lagrange multiplier is assigned to the peak power constraints since they are considered in the generalized KKT conditions as shown later.

According to the generalized KKT conditions for functional optimization [85, 87], if the optimal solution $P_2^\star(t)$ is a regular point, it satisfies the following conditions:

$$\left. \frac{dl}{dP_2(t)} \right|_{P_2(t)=P_2^\star(t)} \begin{cases} = 0, & 0 < P_2^\star(t) < P_S \\ \leq 0, & P_2^\star(t) = 0 \\ \geq 0, & P_2^\star(t) = P_S \end{cases} \tag{4.12}$$

$$\lambda^\star \geq 0 \tag{4.13}$$

$$\lambda^\star \left(\mathbb{E}_t \left\{ P_2^\star(t) \right\} - P_I l_{21} \right) = 0 \tag{4.14}$$

$$\mathbb{E}_t \left\{ P_2^\star(t) \right\} \leq P_I l_{21} \tag{4.15}$$

where the function l is defined as

$$l = \ln \left(1 + \frac{P_2(t)}{t} \right) + \lambda^\star \left(P_2(t) - P_I l_{21} \right) \tag{4.16}$$

and the derivative of (4.12) with respect to (w.r.t.) $P_2(t)$ is

$$\frac{dl}{dP_2(t)} = \frac{1}{t + P_2(t)} - \lambda^\star. \tag{4.17}$$

By solving the KKT conditions (4.12), (4.13), and using $v = 1/\lambda$, we obtain the solution as given in (4.10). The KKT conditions (4.14) and (4.15) are used to determine $v^\star$. Particularly, if the constraint (4.6) is inactive at the optimal solution, $v^\star$ should be

infinity since $\lambda^\star = 0$ according to (4.14). Under this circumstance, the optimal solution is $P_2^\star(t) = P_S$ for all t, which leads to $\mathbb{E}_t\left\{P_2^\star(t)\right\} = P_S < P_I l_{21}$. If $P_S = P_I l_{21}$, it is easy to see that the optimal solution is also $P_2^\star(t) = P_S$, thus the calculation of $v^\star$ is not required. Thus, we only focus on calculating $v^\star$ for $P_S > P_I l_{21}$. In this case, $v^\star$ should be chosen to let the equality $\mathbb{E}_t\left\{P_2^\star(t)\right\} = P_I l_{21}$ hold. □

In order to determine three regions of t to divide the power allocation strategy in (4.10), it is essential to efficiently calculate $v^\star$. As discussed, we only need to obtain $v^\star$ for $P_S > P_I l_{21}$. Defining

$$Q_{avg}(v) = P_S \int_0^{v-P_S} f(t)dt + \int_{v-P_S}^{v} (v-t)f(t)dt \tag{4.18}$$

where $f(t)$ is the PDF of t. We need to choose $v^\star$ to let

$$Q_{avg}(v^\star) = P_I l_{21} \tag{4.19}$$

hold under such circumstance. Taking the derivative of $Q_{avg}(v)$ over v results in

$$Q_{avg}^{'}(v) = \frac{dQ_{avg}(v)}{dv} = \int_{v-P_S}^{v} f(t)dt > 0 \tag{4.20}$$

which shows that $Q_{avg}(v)$ is a strictly monotonical increasing function on v, $\forall\ v > 0$. This property makes some general numerical approaches, i.e., the bisection method, feasible to seek for $v^\star$. However, due to the lack of the explicit expressions of these integrals in (4.18), we need to evaluate v in each bisection step over a large number of channel realizations, which is very time consuming.

We propose an efficient method to determine $v^\star$ that uses the closed-form expressions of $Q_{avg}(v)$ and its derivative $Q_{avg}^{'}(v)$. Specifically, using the Newton's method, we calculate the following iteration until a sufficiently accurate value is reached

$$v^{(n+1)} = v^{(n)} - \frac{Q_{avg}(v^{(n)}) - P_I l_{21}}{Q_{avg}^{'}(v^{(n)})} \tag{4.21}$$

where the closed-form expressions of $Q_{avg}(v)$ and $Q_{avg}^{'}(v)$ are given in Appendix A.2. The initial point of the iteration need to be selected carefully to guarantee the method converges [105].

The overall method seeking for $v^\star$ is as follows. We first choose some initial value of v and apply the Newton's method in (4.21). If it fails to converge after a pre-defined number of iterations, it means that the initial point is not properly chosen. In such cases, an alternative bisection method is applied by using the analytical expression of $Q_{avg}(v)$ in each iteration step.

Remarks: If the interference of the PT-SR link is ignored, then the optimization problem can be solved similarly to the method in [69]. The resulting power allocation strategies are also divided into three regions. Herein, we not only additionally consider this interference term, but also derive the analytical forms of $Q_{avg}(v)$ and

$Q'_{avg}(v)$ in (A.6) and (A.7), respectively. Based on these closed-form expressions, an efficient method is proposed to calculate the threshold value v in determining different power allocation regions.

4.2.1.2 Performance Analysis

We characterize the the achievable performance of a single-antenna secondary link using the optimal power allocation strategy in (4.10).

Proposition 4.2.2. *The achievable performance of exploiting the power allocation strategy (4.10) is given as*

- *If* $P_S \leq P_I l_{21}$, *then*

$$R\left(P_2^\star(t)\right) = \lim_{v^\star \to \infty} r\left(v^\star - P_S, 1, P_S\right) \tag{4.22}$$

- *If* $P_S > P_I l_{21}$, *then*

$$R\left(P_2^\star(t)\right) = r\left(v^\star - P_S, 1, P_S\right) + r\left(v^\star, 0, v^\star\right) - r\left(v^\star - P_S, 0, v^\star\right) \tag{4.23}$$

where the function $r(T, \alpha, \beta)$ *with* $\alpha + \beta/T > 0$. *For* $T \leq 0$, $r(T, \alpha, \beta) = 0$; *for* $T > 0$,

$$r(T,\alpha,\beta) = \begin{cases} \frac{b}{g}\left(\frac{\ln(\alpha+\gamma)}{\gamma+h}e^{-g\gamma} + \frac{1}{\alpha-h}\left(e^{gh}E_1(g(h+\gamma)) - e^{g\alpha}E_1(g(\alpha+\gamma))\right)\right), & h \neq \alpha \\ \frac{b}{g}\left(\frac{\ln(\gamma+\alpha)}{\gamma+\alpha}e^{-g\gamma} + \frac{e^{-g\gamma}}{\gamma+\alpha} - ge^{g\alpha}E_1(g(\gamma+\alpha))\right), & h = \alpha \end{cases} \tag{4.24}$$

In above, we use the variables

$$\gamma = \frac{\beta}{T},\ b = \frac{l_{12}\sigma_S^2}{P_1},\ g = \frac{l_{22}\sigma_S^2}{\beta},\ h = \frac{l_{12}\beta}{l_{22}P_1}. \tag{4.25}$$

Proof. See Appendix A.3. □

Two remarks are addressed here and will be verified in the numerical results.

1. We consider the relation between the achievable rate and the peak transmit power constraint P_S given a certain value of P_I.

   First, when P_S increases from zero and fulfills $P_S < P_I l_{21}$, $v^\star$ is infinite. It is obvious that the derivative of (A.9) over P_S is larger than zero. Therefore, the achievable rate increases monotonically with P_S.

Second, we consider P_S increasing beyond $P_I l_{21}$ but smaller than $\widehat{P_{S,1}}$ (its value will be discussed later). Calculating the derivative of Q_{avg} in (A.6) over P_S,[2] we have

$$\frac{dQ_{avg}}{dP_S} = \int_0^{v-P_S} f(t)dt \geq 0. \tag{4.26}$$

Combining it with (4.20) means that an increasing P_S yields a decreasing $v^\star$ since Q_{avg} is fixed in this case.

Third, keeping P_S increasing and consequently yielding a decreasing $v^\star$, we reach some boundary point $\widehat{P_{S,1}}$ beyond which $v^\star - P_S < 0$ holds. Reviewing (A.9), the achievable rate remains constant even if P_S increases further. Taking $\widehat{P_{S,1}} = \widehat{v^\star}$ into (A.6), the boundary $\widehat{P_{S,1}}$ can be calculated by solving the following equation

$$\int_0^{\widehat{P_{S,1}}} (\widehat{P_{S,1}} - t) f(t)dt = P_I l_{21}. \tag{4.27}$$

2. We consider the relation between the achievable rate and the average interference constraint P_I given a certain value of P_S.

   First, if we increase P_I from zero to $\widehat{P_{I,1}} = P_S/l_{21}$, the mean interference constraint should be met with equality at the optimal value, i.e., $v^\star$ is finite. From (A.7) we observe that $Q_{avg}(v^\star)$ strictly increases with $v^\star$. Thus, an increasing P_I yields an increasing $v^\star$. Taking the derivative of R in (A.9) with regard to v, we note that the achievable rate increases with finite $v^\star$ and P_I.

   Then, if P_I keeps increasing and fulfills $P_I > P_S/l_{21}$, the average interference constraint is met with inequality at the optimal value in (4.8) since the optimal power is always equal to P_S. This indicates that the achievable rate is constant.

The numerical results for the achievable rate of the secondary link with the optimal power control (4.10) are represented as follows. We set $l_{12} = 3$, $l_{21} = 2$ and $l_{22} = 1$. The noise variance of the secondary link is $\sigma_S^2 = 1$. We compare both the numerical and analytical results of the optimal solution to the non-adaptive power transmission aiming at satisfying the system constraints at every time instant:

$$P_2(t) = \begin{cases} P_S, & P_S \leq P_I l_{21} \\ P_I l_{21}, & \text{otherwise} \end{cases}, \quad \forall t > 0. \tag{4.28}$$

Fig. 4.2 plots the achievable rate versus the transmit power constraint P_S under interference power constraints P_I. The detailed description of the legend is given below the figure. The analytical expression of the achievable rate matches exactly the numerical solutions. Furthermore, an increasing P_S with a fixed P_I yields an increasing achievable rate until some boundary. Increasing P_S beyond this boundary

[2] Herein, we take Q_{avg} as the function of P_S instead of v. Therefore, the argument (v) is omitted in $Q_{avg}(v)$.

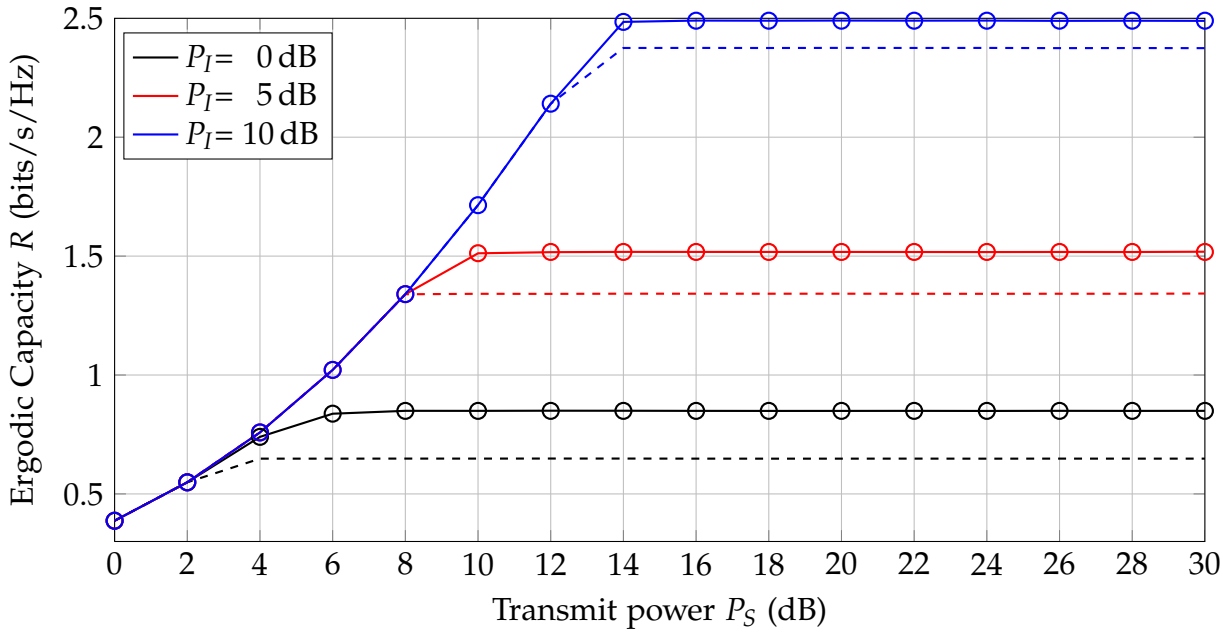

Figure 4.2: Achievable rate versus transmit power limit P_S for different interference power limit P_I. The lines/markers of the same color share the same parameter P_S. Solid lines: optimal solution (sim.); circles: optimal solution I (analytical); dashed lines: non-adaptive power transmission (sim.).

does not affect the achievable rate. The value of the boundary can be obtained by solving (4.27). Finally, the superiority of the optimal solution over the non-adaptive transmission is clearly shown.

Fig. 4.3 illustrates the achievable rate versus the interference power constraint P_I under different transmit power constraints P_S. The optimal solution outperforms the non-adaptive transmission due to the adaptation to the available instantaneous CSI. We observe the rate saturates with the increasing of P_I when P_S is fixed. Given a certain P_S, the boundary P_I is given by $P_S l_{21}^{-1}$. For example, for the case $P_S = 10\,\text{dB}$, the boundary is $\widehat{P_{I,1}} = 6.9897\,\text{dB}$ which matches the observation. When P_I increases beyond $\widehat{P_{I,1}}$, the optimal power allocation turns out to be a fixed power transmission with P_S which consequently yields the constant achievable rate.

4.2.2 Unknown CSI of the PT-SR Link

Different from Section 4.2.1, we assume that the ST knows the statistical CSI of the PT-SR link instead of the instantaneous CSI. More specifically, the ST acquires the instantaneous power gain g_{22} and statistical CSI l_{12} and l_{21}. Two constraints are imposed: the average interference power received at the PR is constrained by P_I and the peak power of the ST is limited by P_S. The optimization problem is formulated as

$$\tilde{R} = \max_{P_2(g_{22})} \mathbb{E}_{g_{12},g_{22}} \left\{ \ln\left(1 + \frac{P_2(g_{22})g_{22}}{P_1 g_{12} + \sigma_S^2}\right) \right\} \tag{4.29}$$

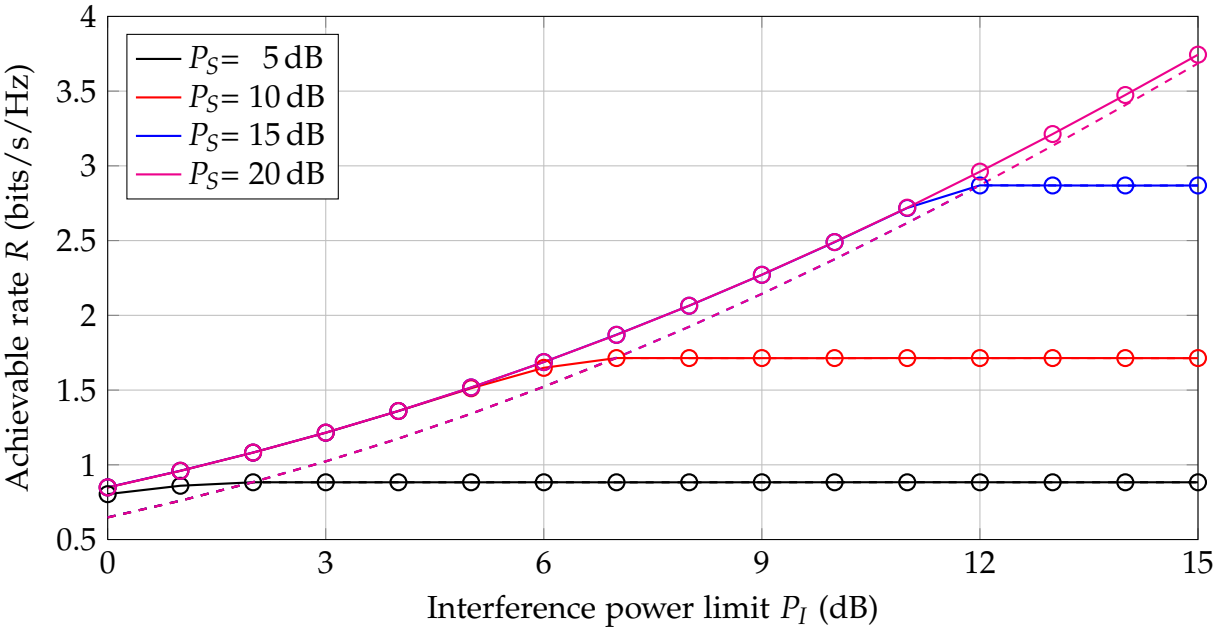

Figure 4.3: Achievable rate versus interference power limit P_I for different transmit power limit P_S. The lines/markers of the same color share the same parameter P_S. Solid lines: optimal solution (sim.); circles: optimal solution I (analytical); dashed lines: non-adaptive power transmission (sim.).

$$\begin{aligned} \text{s.t.} \quad & P_2(g_{22}) \leq P_S, \qquad \forall g_{22} \geq 0 \\ & P_2(g_{22}) \geq 0, \qquad \forall g_{22} \geq 0 \\ & \mathbb{E}_{g_{22},g_{21}} \{P_2(g_{22}) g_{21}\} \leq P_I. \end{aligned}$$

Note that here the notation $P_2(g_{22})$ is used instead of $P_2(t)$ in (4.8) because the secondary transmit power is only adapted to the instantaneous CSI g_{22}. Because the expectation operation preserves convexity [115], the objective function of (4.29) is a concave function in $P_2(g_{22})$. Moreover, all the constraints in (4.29) are affine functions of $P_2(g_{22})$. Hence, (4.29) is a convex optimization problem.

The average interference power constraint in (4.29) is converted into

$$\mathbb{E}_{g_{22}} \{P_2(g_{22})\} \leq \frac{P_I}{\mathbb{E}_{g_{21}} \{g_{21}\}} = P_I l_{21} \tag{4.30}$$

which corresponds to an average transmit power constraint.

Remarks: After reformulating the average interference power constraint into the average power constraint, the problem (4.29) is equivalent to the problem aiming at maximizing the secondary rate subject to both a peak and an average power constraint. The problem formulation of (4.29) generalizes problems studied previously in the literature. If we set $P_1 = 0$, (4.29) reduces to the problems considered in [69] and [71],[3] i.e., without considering the interference from the PT.

[3] In [71], optimal power control in a multi-antenna setup is considered as well.

4.2.2.1 Optimal Power Allocation

We can show that a regularity condition, e.g., the Slater condition, is satisfied. Usually, the optimum of such an optimization problem is mathematical tractable. However, deriving the analytical solution to (4.29) still involves several challenges, e.g., it is hard to obtain an explicit analytical form since the objective function includes an expectation expression.

In what follows, we first reformulate the objective function as an explicit function of the available CSI. Based on this, we solve (4.29) by applying the KKT optimality conditions. The resulting solution requires solving a nonlinear equation. An efficient numerical method is provided to obtain the solution and the thresholds in dividing different power allocation regions.

The power gain of the interfering PT-SR link g_{12} follows an exponential distribution with mean $1/l_{12}$. Moreover, the optimal power P_2 is a function of statistical CSI l_{12} instead of the unknown instantaneous CSI g_{12}. Motivated by this fact, we evaluate the objective function of (4.29) over different channel realizations by calculating the expectation with respect to (w.r.t.) g_{12}. Specifically, introducing

$$s_0 = \frac{\sigma_S^2}{P_1}, \quad s_1 = \frac{\sigma_S^2 + P_2(g_{22})g_{22}}{P_1} \tag{4.31}$$

we write the objective function as

$$\begin{aligned}
&\mathbb{E}_{g_{12},g_{22}}\left\{\ln\left(1+\frac{P_2(g_{22})g_{22}}{\sigma_S^2+P_1g_{12}}\right)\right\}\\
&\overset{(a)}{=} \mathbb{E}_{g_{22}}\left\{\ln\left(1+\frac{P_2(g_{22})g_{22}}{\sigma_S^2}\right) - \frac{P_2(g_{22})g_{22}}{P_1}\int_0^\infty \frac{e^{-l_{12}g_{12}}dg_{12}}{g_{12}^2+\frac{2\sigma_S^2+P_2(g_{22})g_{22}}{P_1}g_{12}+\frac{\sigma_S^4+P_2(g_{22})g_{22}\sigma_S^2}{P_1^2}}\right\}\\
&\overset{(b)}{=} \mathbb{E}_{g_{22}}\left\{\ln\left(1+\frac{P_2(g_{22})g_{22}}{\sigma_S^2}\right) - \int_0^\infty\left(\frac{1}{g_{12}+s_0}-\frac{1}{g_{12}+s_1}\right)e^{-l_{12}g_{12}}dg_{12}\right\}\\
&\overset{(c)}{=} \mathbb{E}_{g_{22}}\left\{\ln\left(1+\frac{P_2(g_{22})g_{22}}{\sigma_S^2}\right) + e^{l_{12}s_1}E_1(l_{12}s_1)\right\} - e^{l_{12}s_0}E_1(l_{12}s_0)
\end{aligned} \tag{4.32}$$

where, in step (a), we use integration by parts and incorporating the PDF of g_{21}, and in step (b) we integrating (4.31). In step (c), we use the definition of the exponential integral [4]

$$E_1(x) = \int_x^\infty \frac{e^{-t}}{t}dt, \quad \mathcal{R}(x) > 0. \tag{4.33}$$

Since (4.29) is a convex optimization problem with continuous differentiable objective function and constraints and the Slater condition is satisfied, the KKT conditions are sufficient and necessary conditions for the optimum solution [87].

Theorem 4.2.1. *The optimal power allocation $P_2^\star(g_{22})$ solving the problem (4.29) is*

$$P_2^\star(g_{22}) = \begin{cases} 0, & g_{22} \le v_0(\lambda_2^\star) \\ p^\star(g_{22}), & v_0(\lambda_2^\star) < g_{22} < v_1(\lambda_2^\star) \\ P_S, & \textit{otherwise}. \end{cases} \tag{4.34}$$

where $\lambda_2^\star$ is the optimal value of the non-negative Lagrangian multiplier corresponding to the constraint (4.30). Here, $v_0(\lambda_2^\star)$ and $v_1(\lambda_2^\star)$ are the thresholds that divide different power level regions. We have $v_0(\lambda_2^\star) = \lambda_2^\star/(be^a E_1(a))$ and $H(v_1(\lambda_2^\star)) = \lambda_2^\star$ with

$$H(x) = b_1 x e^{a_1 + b_1 P_S x} E_1\left(a_1 + b_1 P_S x\right) \tag{4.35}$$

and

$$a_1 = \frac{l_{12}\sigma_S^2}{P_1}, \quad b_1 = \frac{l_{12}}{P_1}. \tag{4.36}$$

The optimal power value in the intermediate region is

$$p^\star(g_{22}) = \frac{G^{-1}\left(\lambda_2^\star/(b_1 g_{22})\right) - a_1}{b_1 g_{22}} \tag{4.37}$$

with

$$G(x) = e^x E_1(x), \quad \forall\, x > 0. \tag{4.38}$$

Proof. Regarding the reformulated optimization problem of (4.29) with objective function (4.32).[4] The Lagrangian of (4.29) is

$$\begin{aligned} L = \mathbb{E}_{g_{22}} &\left\{ \ln\left(1 + \frac{P_2(g_{22}) g_{22}}{\sigma_S^2}\right) + e^{l_{12}s_1} E_1(l_{12}s_1) \right\} \\ &- \lambda_2 \left[\mathbb{E}_{g_{22}} \left\{ P_2(g_{22}) \right\} - P_I l_{21} \right]. \end{aligned} \tag{4.39}$$

where the non-negative Lagrangian multiplier λ_2 corresponds to the constraint (4.30). Herein, no Lagrange multipliers are assigned to the peak power constraints since they are considered in the generalized KKT conditions.

According to the generalized KKT conditions for functional optimization [85, 87], the optimal solution $P_2^\star(g_{22})$ satisfies the following conditions:

$$\left.\frac{dl}{dP_2(g_{22})}\right|_{P_2(g_{22}) = P_2^\star(g_{22})} \begin{cases} \ge 0, & P_2^\star(g_{22}) = P_S \\ \le 0, & P_2^\star(g_{22}) = 0 \\ = 0, & 0 < P_2^\star(g_{22}) < P_S \end{cases} \tag{4.40}$$

$$\lambda_2^\star \ge 0 \tag{4.41}$$

$$\lambda_2^\star \left(\mathbb{E}_{g_{22}} \left\{ P_2(g_{22}) \right\} - P_I l_{21} \right) = 0 \tag{4.42}$$

$$\mathbb{E}_{g_{22}} \left\{ P_2(g_{22}) \right\} \le P_I l_{21} \tag{4.43}$$

[4] The term $e^{l_{12}s_0} E_1(l_{12}s_0)$ in (4.32) is independent of $P_2(g_{22})$. Therefore, we omit it in the algorithmic design since it does not affect the final result.

where the function l is defined as

$$l = \ln\left(1 + \frac{P_2(g_{22})g_{22}}{\sigma_S^2}\right) + e^{l_{12}s_1}E_1(l_{12}s_1) - \lambda_2^\star\left(P_2(g_{22}) - P_I l_{21}\right) \tag{4.44}$$

and the derivative of (4.40) w.r.t. $P_2(g_{22})$ is

$$\frac{dl}{dP_2(g_{22})} = b_1 g_{22} e^{a_1 + b_1 g_{22} P_2^\star(g_{22})} E_1(a_1 + b_1 g_{22} P_2^\star(g_{22})) - \lambda_2^\star \tag{4.45}$$

where a_1 and b_1 are given in (4.36). Solving the KKT conditions, we obtain the optimal solution $P_2^\star(g_{22})$. It is divided into three cases according to different values of g_{22}:

1. $P_2^\star(g_{22}) = P_S$. According to the first inequality in (4.40), this holds under the condition
$$b_1 g_{22} e^{a_1 + b_1 P_S g_{22}} E_1(a_1 + b_1 P_S g_{22}) \geq \lambda_2^\star. \tag{4.46}$$

2. $P_2^\star(g_{22}) = 0$. According to the second inequality in (4.40), this holds under the condition
$$b_1 e^{a_1} E_1(a_1) g_{22} \leq \lambda_2^\star. \tag{4.47}$$

3. $0 < P_2^\star(g_{22}) < P_S$. According to the third equality in (4.40), we have the solution $P_2^\star(g_{22}) = p^\star(g_{22})$ that satisfies the following nonlinear equation
$$b_1 g_{22} e^{a_1 + b_1 g_{22} p^\star(g_{22})} E_1(a_1 + b_1 g_{22} p^\star(g_{22})) = \lambda_2^\star. \tag{4.48}$$

As seen from (4.46), (4.47), and (4.48), $P_2^\star(g_{22} = 0) = 0$ holds if $\lambda_2^\star$ is positive. For the special case $\lambda_2^\star = 0$, $P_2^\star(g_{22} = 0)$ can be chosen as an arbitrary value in $[0, P_S]$ and does not affect the objective function. Herein, we choose the solution $P_2^\star(g_{22} = 0) = 0$ for $\lambda_2^\star = 0$ to be consistent with the former case. In what follows, we only focus on the discussion of the optimal power value $P_2^\star(g_{22})$ for positive g_{22}.

Using the definition of $G(x)$ in (4.36), the equation (4.48) is reformulated into

$$b_1 g_{22} G(a_1 + b_1 g_{22} p^\star(g_{22})) = \lambda_2^\star \tag{4.49}$$

We prove in Appendix A.4 that $G(x)$ is strictly monotonically decreasing in x for positive and finite x. Consequently, the optimal solution of $p^\star(g_{22})$ in (4.48) is given in (4.37).

The solution of (4.37) is considered feasible if it is between zero and P_S. Considering that $G(a_1 + b_1 g_{22} x)$ is also a monotonically decreasing function with x, $x > 0$, for positive a_1, b_1, and g_{22}, we have the feasibility solutions (4.49) in the region $0 < p^\star(g_{22}) < P_S$ as

$$b_1 g_{22} G(a_1 + b_1 g_{22} P_S) < \lambda_2^\star < b_1 g_{22} G(a_1). \tag{4.50}$$

The effect in the variations of g_{22} on the optimal power $P_2^\star(g_{22})$ is not explicitly visible in (4.46), (4.47), and (4.50). In what follows, we aim at addressing this issue.

First, the left side of the first condition (4.46) is a nonlinear function in g_{22}. We show that $H(x)$ in (4.35) is a monotonically increasing function for $x \geq 0$ and $\lim_{x\to 0} H(x) = 0$ in Appendix A.5. Therefore, in order to satisfy (4.46), we require that $g_{22} \geq v_1(\lambda_2^\star)$, where $v_1(\lambda_2^\star)$ is a lower bound on g_{22}. The value of $v_1(\lambda_2^\star)$ is equal to the root of $H(x) = \lambda_2^\star$ and can be evaluated numerically. For the special case $\lambda_2^\star = 0$, we have $v_1(\lambda_2^\star) = 0$. Considering the second condition (4.47) for $P_2^\star(g_{22}) = 0$, we rewrite it as $g_{22} \leq v_0(\lambda_2^\star)$ with $v_0(\lambda_2^\star) = \lambda_2^\star/(b_1 e^{a_1} E_1(a_1))$. This indicates that if the power gain of the ST-SR link is smaller than $v_0(\lambda_2^\star)$, the optimal strategy is to switch off the ST. Concerning the third case with $0 < P_2^\star(g_{22}) < P_S$, where the power level $P_2^\star(g_{22})$ is given in (4.37), the condition (4.50) defines the complementary region given by the first condition (4.46) and the second condition (4.47). Therefore, it corresponds to the case

$$v_0(\lambda_2^\star) < g_{22} < v_1(\lambda_2^\star). \tag{4.51}$$

For comparison, we apply the KKT conditions of functional optimization [87] to the special case in [69,71] that equivalent to $P_1 = 0$. The same result is obtained. □

Two questions arise from the above derived solution (4.34).

1. What is an efficient numerical method to obtain $p^\star(g_{22})$ from (4.37)?
2. How to determine the parameter $\lambda_2^\star$ in (4.34)?

In the following, we focus on answering these two problems.

The Algorithm for Solving (4.37) Eq. (4.37) is rewritten into

$$b_1 g_{22} G(a_1 + b_1 g_{22} p^\star(g_{22})) = \lambda_2^\star \tag{4.52}$$

According to (4.34), a feasible condition of $p^\star(g_{22}) \in (0, P_S)$ is

$$v_0(\lambda_2^\star) < g_{22} < v_1(\lambda_2^\star).$$

Moreover, $G(a_1 + b_1 g_{22} x)$ is strictly monotonically decreasing in x for positive and finite a_1, b_1 and g_{22}. Combined with these two factors, the value of $p^\star(g_{22})$ to solve (4.52) can be calculated by the bisection method. The bisection method repeatedly bisects an interval and selects a subinterval in which an optimum must lie for further processing until convergence. It has a linear rate of convergence [19].

Determination of $\lambda_2^\star$ in (4.34) We discuss this issue under two circumstances. On the one hand, if $P_S \leq P_I l_{21}$, the ST can always transmit with power P_S while the average power constraint (4.30) is strictly satisfied, i.e., the optimal solution is constant and equal to P_S and the computation of $\lambda_2^\star$ is not required. On the other hand, if $P_S > P_I l_{21}$, using $f_{g_{22}}(x)$ as the probability density function of g_{22}, we have

$$Q_{\text{avg}} = \mathbb{E}_{g_{22}}\left\{P_2^\star(g_{22})\right\} = \int_{v_0(\lambda_2^\star)}^{v_1(\lambda_2^\star)} p^\star(x) f_{g_{22}}(x) dx + P_S \int_{v_1(\lambda_2^\star)}^{\infty} f_{g_{22}}(x) dx. \tag{4.53}$$

The optimal $\lambda_2^\star$ should be chosen to satisfy $Q_{\text{avg}} = P_I l_{21}$. We show that $\lambda_2^\star$ is finite and positive in this case. The argument is proved by reduction to absurdity: If $\lambda_2^\star = 0$, it follows that $v_0(\lambda_2^\star) = v_1(\lambda_2^\star) = 0$. Integrating this into (4.34), we have $P_2^\star(g_{22}) = P_S$ for any positive g_{22}. Consequently, this yields $Q_{\text{avg}} = P_S > P_I l_{21}$ which violates (4.30). Similarly, we have $P_2^\star(g_{22}) = 0$ and $Q_{\text{avg}} = 0$ if $\lambda_2^\star$ is infinitely large, which is obviously not optimal unless we consider the trivial scenario under the condition $P_I > 0$. Therefore, $\lambda_2^\star$ is finite and positive.

In order to obtain $\lambda_2^\star$ to satisfy the equation (4.30) for the case $P_S > P_I l_{21}$, we calculate the derivative of Q_{avg} over $\lambda_2^\star$ and obtain

$$\frac{dQ_{\text{avg}}}{d\lambda_2^\star} = \int_{v_0(\lambda_2^\star)}^{v_1(\lambda_2^\star)} \frac{dp^\star(x)}{d\lambda_2^\star} f_{g_{22}}(x)dx \overset{(a)}{<} 0 \tag{4.54}$$

where the inequality (a) uses the property that the optimal solution decreases as $\lambda_2^\star$ increases. Given the strict monotonicity of Q_{avg} over $\lambda_2^\star$ in (4.54), the value of $\lambda_2^\star$ can be calculated using the bisection method that has a linear rate of convergence [19].

Finally, we investigate the relation between the achievable performance and the system constraints by using the optimal solution (4.34). We show that the achievable rate remains constant provided either P_S or P_I is fixed and increasing P_I or P_S beyond some limit given by $\widehat{P_{I,2}}$ and $\widehat{P_{S,2}}$, respectively.

On the one hand, if P_I keeps increasing and satisfies $P_I \geq P_S/l_{21}$, the optimal power is equal to P_S which results in a constant achievable rate $\tilde{R}$ in this case. Therefore, we have $\widehat{P_I} = P_S/l_{21}$. On the other hand, we have

$$b_1 g_{22} e^{a_1 + b_1 P_S g_{22}} E_1(a_1 + b_1 P_S g_{22}) < b_1 g_{22}/(a_1 + b_1 P_S g_{22}) < 1/P_S \tag{4.55}$$

by applying (A.15) in Appendix A.4 to the left side of (4.46). This means that given a certain $\lambda_2^\star$, increasing P_S beyond some boundary point $\widehat{P_{S,2}}$, i.e., the constraint (4.46) is always violated. The optimal power never reaches the peak power limit P_S. Therefore, increasing P_S does not affect the achievable rate. The threshold $\widehat{P_{S,2}}$ can be approximated by $\bar{P}_S$ that is obtained from the following equality:

$$\int_{\bar{v}_0 = 1/(b_1 e^{a_1} G(a_1) \bar{P}_S)}^{\infty} p^\star(x) f_{g_{22}}(x)dx = P_I l_{21}. \tag{4.56}$$

4.2.2.2 Suboptimal Power Allocation I

The optimal power allocation strategy in (4.34) requires two numerical computation steps: the computation of $p^\star(g_{22})$ and the calculation of $\lambda_2^\star$. The value of $p^\star(g_{22})$ needs to be calculated for every time instant, while the value of $\lambda_2^\star$ needs to be updated with changing statistical channel parameters. Hence, calculating the optimal solution is time-consuming. Moreover, due to the lack of the optimal solution in closed form, deriving the analytical performance is intractable. In order to reduce the algorithm complexity, we propose two suboptimal low complexity strategies in Section 4.2.2.2 and Section 4.2.2.3. The first one, named double threshold waterfilling

(DT-WF), is based on an approximation of the optimal solution. The second strategy, named double threshold constant-power waterfilling (DTCP-WF), further simplifies DT-WF. Additionally, benefited from the analytical expression of the solutions, the performance of two suboptimal solutions is given in closed form.

The real-time computation of $p^\star(g_{22})$ requires numerically solving an equation in the form of

$$G(x) = C$$

where C is some positive constant. In order to obtain a low-complexity strategy, we resort to an approximate solution of x. Particularly, we use both the upper-bound and the lower-bound of $G(x)$ in (A.15) as two approximations

$$G(x) \approx \frac{1}{x} \tag{4.57}$$

$$G(x) \approx \frac{1}{x+1} \tag{4.58}$$

respectively. By checking the approximation accuracy of both expressions, we observe that both approximations are close to $G(x)$ when x is large. Among the two of them, the approximation $G(x) \approx 1/(x+1)$ is close to $G(x)$ in general. Therefore, we apply $G(x) \approx 1/(x+1)$ in the remainder of this section. Specifically, $p^\star(g_{22})$ in (4.37) is approximated by

$$\tilde{p}^\star(g_{22}) = v_2^\star - \frac{a_1+1}{b_1 g_{22}} \tag{4.59}$$

where the calculation of $v_2^\star$ will be addressed later.

Similarly, we use $G(x) \approx 1/(x+1)$ to approximate the conditions in (4.34) under which $P_2^\star(g_{22})$ is equal to $0, p^\star(g_{22}), P_S$, respectively. After some mathematical manipulations, the suboptimal power control strategy DT-WF is obtained as

$$v_2^\star > P_S: \quad \tilde{P}_2^\star(g_{22}) = \begin{cases} 0, & g_{22} \le \dfrac{a_1+1}{b_1 v_2^\star} \\ \tilde{p}^\star(g_{22}), & \dfrac{a_1+1}{b_1 v_2^\star} < g_{22} < \dfrac{a_1+1}{b_1\left(v_2^\star - P_S\right)} \\ P_S, & \text{otherwise} \end{cases} \tag{4.60}$$

$$0 < v_2^\star \le P_S: \quad \tilde{P}_2^\star(g_{22}) = \begin{cases} 0, & g_{22} \le \dfrac{a_1+1}{b_1 v_2^\star} \\ \tilde{p}^\star(g_{22}), & \text{otherwise.} \end{cases} \tag{4.61}$$

One crucial question in calculating $\tilde{P}_2^\star(g_{22})$ is how to determine $v_2^\star$ in (4.60) and (4.61). If $P_S \le P_I l_{21}$, the suboptimal power is $\tilde{P}_2^\star(g_{22}) = P_S$. Hence, the computation of $v_2^\star$ is not required. Otherwise in the case $P_S > P_I l_{21}$, $v_2^\star$ is finite and positive. In the following, we propose an efficient method to calculate $v_2^\star$.

Proposition 4.2.3. *For $P_S > P_I l_{21}$, the positive parameter $v_2^\star$ is uniquely chosen to satisfy*

$$F_2\left(v_2^{(\star)}\right) = P_I l_{21} \tag{4.62}$$

with $F_2(x)$ given by

$$F_2(x) = f_2(x) - f_2(x - P_S) \tag{4.63}$$

and the definition of $f_2(x)$ as

$$f_2(x) = \begin{cases} xe^{-\frac{\lambda_b}{x}} - \lambda_b E_1\left(\frac{\lambda_b}{x}\right), & x > 0 \\ 0, & \text{otherwise.} \end{cases} \tag{4.64}$$

with $\lambda_b = (a_1 + 1)l_{22}/b_1$. The value of $v_2^\star$ can be obtained by applying the Newton's method

$$v_2^{(n+1)} = v_2^{(n)} - \frac{F_2\left(v_2^{(n)}\right) - P_I l_{21}}{F_2'(v_2^{(n)})} \tag{4.65}$$

after convergence. The sequence (4.65) satisfies the standard assumptions of the local convergence theorem for the Newton's method [67].

The initial point of the iteration is provided by

$$v_2^{(0)} = -\frac{(a_1 + 1)l_{22}}{b_1 \ln\left(P_I l_{21}/P_S\right)}. \tag{4.66}$$

Proof. See Appendix A.6. □

A theoretical performance analysis is desirable in designing the algorithm since the performance can be evaluated without running numerous simulations. Considering the suboptimal solution (4.60) and (4.61) in which the explicit connection between $\tilde{P}_2^\star(g_{22})$ and the instantaneous CSI g_{22} is shown, deriving an analytical performance result is feasible. The main result is summarized as follows.

Proposition 4.2.4. *Given the functions $\tilde{r}_1(\lambda, p)$ and $\tilde{r}_2(\lambda)$ in (A.31) and (A.39) and*

$$g_1 = \frac{(a_1 + 1)P_1}{(v_2^\star - P_S)l_{12}}, \quad v_2^\star > P_S \tag{4.67}$$

$$g_0 = \frac{(a_1 + 1)P_1}{v_2^\star l_{12}}, \qquad v_2^\star > 0 \tag{4.68}$$

the achievable rate of the suboptimal strategy DT-WF is

$$\tilde{R}_{DT\text{-}WF} = \begin{cases} \tilde{r}_1(g_1, P_S) - \tilde{r}_2(g_1) + \tilde{r}_2(g_0), & v_2^\star > P_S \\ \tilde{r}_2(g_0), & 0 < v_2^\star \leq P_S. \end{cases} \tag{4.69}$$

Proof. See Appendix A.7. □

4.2.2.3 Suboptimal Power Allocation II

In DT-WF, the power level $\tilde{p}^\star(g_{22})$ is calculated in real-time, since it depends on the instantaneous CSI g_{22} in (4.59). In order to further reduce the computational complexity,

we exploit non adaptive power transmission in the region where $\tilde{p}^{\star}(g_{22})$ is allocated. The idea is motivated by the conventional constant-power waterfilling strategy [138] where the transmitter allocates constant power to the channels with nonzero power by optimal waterfilling and zero power in the remaining channels. However, different from the conventional idea, DTCP-WF is designed based on the thresholds of DT-WF instead of the optimal solution since the computation of the thresholds in the optimal solution is more time-consuming.

In DTCP-WF, zero and the peak power value P_S are allocated in the same regions as in DT-WF, otherwise constant power is allocated. Consequently, the form of DTCP-WF is the same as (4.60) and (4.61) except that $\tilde{p}^{\star}(g_{22})$ is replaced by the constant power level $\tilde{p}_c^{\star}$.

The calculation of $\tilde{p}_c^{\star}$ is performed as follows. For the case $P_S \leq P_I l_{21}$, the optimal solution is reduced to transmitting with power P_S in every time instant, thus the calculation of $\tilde{p}_c^{\star}$ is not required. For the case $P_S > P_I l_{21}$, the constant power level $\tilde{p}_c^{\star}(v_2^{\star})$ is chosen to satisfy the average power constraint (4.30). After some mathematical manipulations, we obtain

$$v_2^{\star} > P_S, \quad \tilde{P}_2^{\star}(g_{22}) = \begin{cases} 0, & b_1 g_{22} \leq \dfrac{a_1+1}{v_2^{\star}} \\ P_S, & b_1 g_{22} \geq \dfrac{a_1+1}{v_2^{\star}-P_S} \\ \tilde{p}_c^{\star}(v_2^{\star}), & \text{otherwise} \end{cases}. \tag{4.70}$$

$$0 < v_2^{\star} \leq P_S, \quad \tilde{P}_2^{\star}(g_{22}) = \begin{cases} 0, & b_1 g_{22} \leq \dfrac{a_1+1}{v_2^{\star}} \\ \tilde{p}_c^{\star}(v_2^{\star}), & \text{otherwise} \end{cases}. \tag{4.71}$$

where

$$\tilde{p}_c^{\star} = \begin{cases} \dfrac{P_I l_{21} - P_S e^{-\frac{(a_1+1)l_{22}}{b_1\left(v_2^{\star}-P_S\right)}}}{e^{-\frac{(a_1+1)l_{22}}{b_1 v_2^{\star}}} - e^{-\frac{(a_1+1)l_{22}}{b_1\left(v_2^{\star}-P_S\right)}}}, & v_2^{\star} > P_S \\ \dfrac{P_I l_{21}}{e^{-\frac{(a_1+1)l_{22}}{b_1 v_2^{\star}}}}, & 0 < v_2^{\star} \leq P_S. \end{cases} \tag{4.72}$$

The performance of DTCP-WF can be derived similarly as for DT-WF using the expression $\tilde{r}_1(\lambda, p)$ given in (A.31). For completeness, we directly provide the result:

$$\tilde{R}_{\text{DTCP-WF}} = \begin{cases} \tilde{r}_1(g_1, P_S) - \tilde{r}_1(g_1, \tilde{p}_c^{\star}) + \tilde{r}_1(g_0, \tilde{p}_c^{\star}), & v_2^{\star} > P_S \\ \tilde{r}_1(g_0, \tilde{p}_c^{\star}), & 0 < v_2^{\star} \leq P_S \end{cases} \tag{4.73}$$

where g_1 and g_0 are given in (4.67) and (4.68), respectively.

4.2.2.4 Multiple Primary Links Scenario

The difficulty in solving the considered optimization problem (4.29) arises from the partial CSI related to the primary links. The existence of multiple primary links in the spectrum-sharing system brings more challenges to the optimization problem. In this section, we discuss the extension to the system containing multiple primary links and show that the optimal power allocation strategy can be similarly derived for the general case.

Specifically, we assume there are M PT-PR links coexisting with one ST-SR link in the system. The mth PT transmits at the power level $P_{1,m}$, $\forall\, m = 1, \ldots, M$. The CSI from the mth PT to the SR is denoted by $g_{12,m}$ with its mean $1/l_{12,m}$. Similarly, the CSI from the ST to the mth PR link is $g_{21,m}$ with its mean $1/l_{21,m}$. The primary messages are independent. Consistent with the CSI assumption in the single primary link scenario, the ST is assumed to know the instantaneous CSI of g_{22} and statistical CSI $l_{12,m}$ and $l_{21,m}$, $\forall\, m = 1, \ldots, M$. All channels are assumed to be stationary, ergodic, and mutually independent.

We first define a new variable g_2 denoting the sum of the interference power from M PTs at the SR as

$$g_2 = \sum_{m=1}^{M} P_{1,m} g_{12,m} \tag{4.74}$$

where $P_{1,m} g_{12,m}$, $\forall\, m = 1, \ldots, M$, is independent exponential random variable with expected value given as

$$\frac{1}{\bar{l}_m} = \frac{P_{1,m}}{l_{12,m}}. \tag{4.75}$$

In a practical system, the values of $\bar{l}_m$, $\forall\, m = 1, \ldots, M$, concerning M primary links are distinct. Therefore, for distinct $\bar{l}_m$, $\forall\, m = 1, \ldots, M$, the PDF of g_2 is [72, Eq. (3)]

$$f_{g_2}(g_2) = \sum_{m=1}^{M} A_m \bar{l}_m e^{-\bar{l}_m g_2} \tag{4.76}$$

with

$$A_m = \prod_{u=1, u\neq m}^{M} \frac{1}{1 - \bar{l}_m / \bar{l}_u}. \tag{4.77}$$

Using the definition of g_2 in (4.74), the optimization problem is reformulated as

$$\begin{aligned} (\mathbf{P2}) \quad \max_{P_2(g_{22})} \quad & \mathbb{E}_{g_2, g_{22}} \left\{ \ln \left(1 + \frac{P_2(g_{22}) g_{22}}{g_2 + \sigma_S^2} \right) \right\} \\ \text{s.t.} \quad & P_2(g_{22}) \leq P_S, \qquad \forall g_{22} \geq 0 \\ & P_2(g_{22}) \geq 0, \qquad \forall g_{22} \geq 0 \\ & \mathbb{E}_{g_{21,m}, g_{22}} \{ P_2(g_{22}) g_{21,m} \} \leq P_{I,m} \\ & \forall\, m = 1, \ldots, M \end{aligned} \tag{4.78}$$

where $P_{I,m}$ is the interference power limit of the mth PR. We can show that (4.78) is a convex optimization problem.

Using the independence between the CSI of different links, the last constraints in (4.78) can be simplified to a single constraint

$$\mathbb{E}_{g_{22}}\left\{P_2(g_{22})\right\} \leq P_{I,\mathrm{MU}} \tag{4.79}$$

where $P_{I,\mathrm{MU}} = P_{I,\hat{m}} l_{21,\hat{m}}$ holds and the index $\hat{m}$ is selected according to

$$\hat{m} = \underset{m\in\{1,\ldots,M\}}{\arg\min}\ \left\{P_{I,m} l_{21,m}\right\} \tag{4.80}$$

In order to derive the optimal power allocation strategy, we need to reformulate the objective function in (4.78) as an explicit function of the available CSI. Specifically, inserting (4.76) into the objective function of (4.78)), we have

$$\begin{aligned}
&\mathbb{E}_{g_1, g_{22}}\left\{\ln\left(1+\frac{P_2(g_{22})g_{22}}{\sigma_S^2+g_2}\right)\right\} \\
&= \mathbb{E}_{g_{22}}\Bigg\{\sum_{m=1}^{M} A_m\Bigg(\ln\left(1+\frac{P_2(g_{22})g_{22}}{\sigma_S^2}\right) + e^{\bar{l}_m\left(\sigma_S^2+P_2(g_{22})g_{22}\right)} E_1\left(\bar{l}_m\left(\sigma_S^2+P_2(g_{22})g_{22}\right)\right) \\
&\quad - e^{\bar{l}_m\sigma_S^2} E_1\left(\bar{l}_m\sigma_S^2\right)\Bigg)\Bigg\}.
\end{aligned} \tag{4.81}$$

We integrate (4.79) and (4.81) into (4.78). Applying functional optimization theory [85, 87], and after some mathematical manipulations, we obtain the optimal solution

$$P_2^{\star}(g_{22}) = \begin{cases} 0, & H_1(0) \leq \lambda_{2,\mathrm{MU}}^{\star} \\ p_{\mathrm{MU}}^{\star}(g_{22}), & H_1(P_S) < \lambda_{2,\mathrm{MU}}^{\star} < H_1(0) \\ P_S, & \text{otherwise} \end{cases} \tag{4.82}$$

where the function $H_1(x)$ is defined as

$$H_1(x) = \sum_{m=1}^{M} A_m \bar{l}_m g_{22} e^{\bar{l}_m\left(\sigma_S^2+xg_{22}\right)} E_1\left(\bar{l}_m\left(\sigma_S^2+xg_{22}\right)\right) \tag{4.83}$$

and $p_{\mathrm{MU}}^{\star}(g_{22})$ is the root of the equation

$$H_1(x) = \lambda_{2,\mathrm{MU}}^{\star} \tag{4.84}$$

with $\lambda_{2,\mathrm{MU}}^{\star}$ as the optimal value of the non-negative Lagrangian multiplier that corresponds to the constraint (4.79).

In order to solve $p^{\star}_{\mathrm{MU}}(g_{22})$ from (4.84) efficiently, we check the derivative of $H_1(x)$ w.r.t. x, $x \in (0,\, P_S)$, within its feasibility region given in (4.82). Specifically, we have

$$H_1^{'}(x) = \frac{\partial^2 \mathbb{E}_{g_2}\left\{\ln\left(1 + x g_{22}/(\sigma_S^2 + g_2)\right)\right\}}{\partial x^2} = -\mathbb{E}_{g_2}\left\{\frac{1}{\left((\sigma_S^2 + g_2)/g_{22} + x\right)^2}\right\} < 0 \tag{4.85}$$

for any positive x. Hence, $H_1(x)$ is a strictly monotonically decreasing function in x in the feasible region. Thus, $p^{\star}_{\mathrm{MU}}(g_{22})$ can be obtained from (4.84) using simple numerical methods, e.g., the bisection method.

4.2.2.5 Performance Comparison

We assess the achievable rate in bits/s/Hz. The channel power gain of each fading link g_{ij}, $i, j = 1, 2$ follows an exponential distribution with $l_{12} = 2$, $l_{21} = 3$, and $l_{22} = 1$. The noise variance of the SU is $\sigma_S^2 = 1$. All power values in dB are given relative to the reference level of value 1. The transmit power of the PT is 10dB.

The performance of five power allocation strategies is compared.

- *Optimal algorithm*: The power level value is mapped from g_{22} as given in (4.34).
- *DT-WF*: The power level value is a function of g_{22} as given in (4.60) and (4.61). The analytical performance is provided by (4.69).
- *DTCP-WF*: The power level is matched to g_{22} as given in (4.70) and (4.71), where $\tilde{p}^{\star}(g_{22}, v_2^{\star})$ from DT-WF is replaced by $\tilde{p}_c^{\star}$ in (4.72). The analytical performance is provided by (4.73).
- *Reference algorithm 1*: The first reference algorithm was developed in [69,71]. It can be considered as the result of the special case of (4.29) with $P_1 = 0$, i.e., the interference from the PT is ignored.
- *Reference algorithm 2*: The second reference algorithm is a non-adaptive strategy aiming at satisfying the system constraints at every time instant:

$$P_2(g_{22}) = \begin{cases} P_S, & P_S \leq P_I l_{21} \\ P_I l_{21}, & \text{otherwise.} \end{cases} \tag{4.86}$$

Figure 4.4 plots the achievable rate R versus the peak transmit power constraint P_S. The left hand side subfigure shows the performance under different interference power constraints P_I. The detailed explanations of the legend are given below the figure. We have several observations. Firstly, concerning the achievable performance, the analytical results match the numerical results for the suboptimal solutions. Secondly, the near-optimality of the suboptimal solutions is validated. The gain over the reference non-adaptive power transmission is large for each case of P_I starting

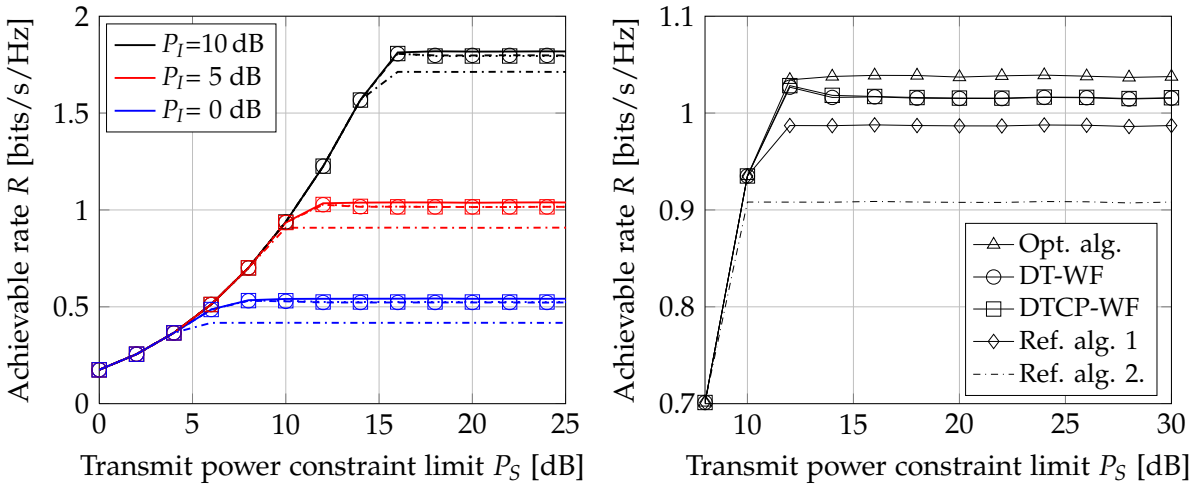

Figure 4.4: Achievable rate of the SU vs. peak transmit power constraint limit P_S with $P_1 = 15$ dB. The lines/markers of the same color share the same setup parameter P_I. Solid lines: optimal solution; dashed lines: DT-WF (sim.); circles: DT-WF (analytical); dotted lines: DTCP-WF (sim.); squares: DTCP-WF (analytical); diamond: ref. alg. 1; dash-dotted lines: ref. alg. 2. In the right hand side subfigure, $P_I = 5$ dB holds.

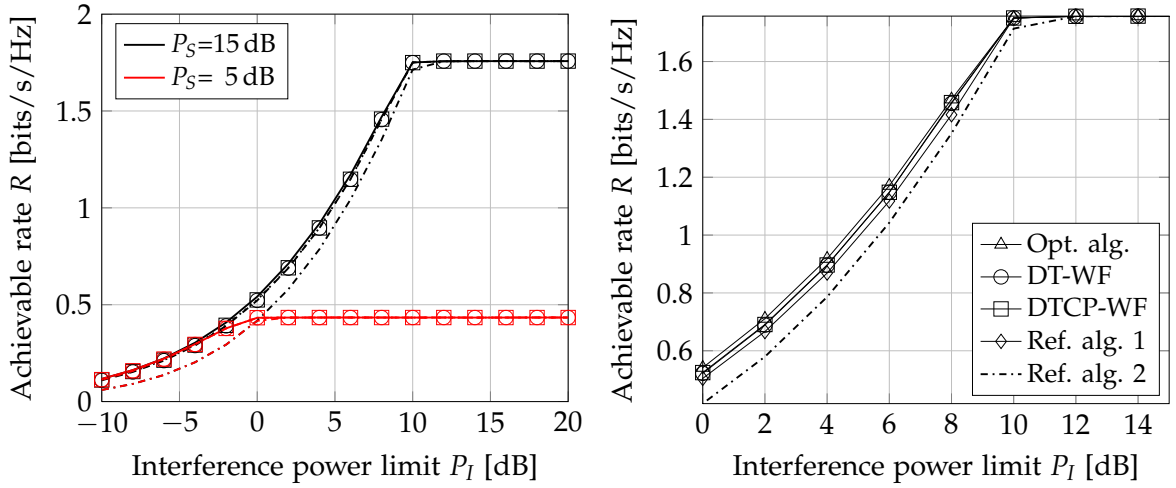

Figure 4.5: Achievable rate of the SU vs. peak transmit power constraint limit P_I with $P_1 = 15$ dB. The lines/markers of the same color share the same setup parameter P_S. Solid lines: optimal solution; dashed lines: DT-WF (sim.); circles: DT-WF (analytical); dotted lines: DTCP-WF (sim.); squares: DTCP-WF (analytical); diamonds: ref. alg. 1; dash-dotted lines: ref. alg. 2. In the right hand side subfigure, $P_S = 15$ dB holds.

at certain P_S. Finally, R increases with an increase of P_S up to some boundary, beyond which it remains constant. The boundary can be calculated with (4.56). The detailed comparison on the performance are shown in the right hand side subfigure

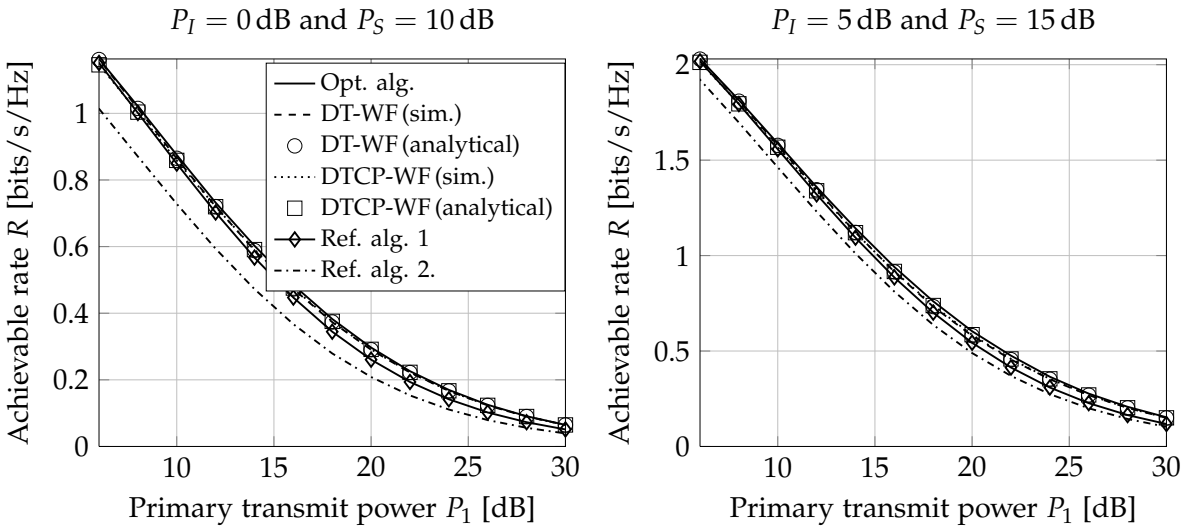

Figure 4.6: Achievable rate of the SU vs. primary transmit power P_1.

for the case of $P_I = 5$ dB. The proposed optimal and suboptimal strategies outperform the reference ones especially for large values of P_S. Considering that our algorithms require no additional instantaneous CSI compared to the reference algorithms, the gain is mainly due to taking the interference term P_1 into account. Besides, we note that the performance loss of the suboptimal strategies compared to the optimal ones increases slightly from 12 dB to 14 dB. The reason is that the approximation (4.59) incurs larger approximation errors for large values of P_S, consequently yielding a larger performance gap between the proposed strategies and the optimal one.

Figure 4.5 depicts the achievable secondary rate R versus the interference power constraint P_I. In the left hand side, the performance is shown under different transmit power constraints P_S. The simulation and analytical results for the suboptimal solutions match. The near-optimality of the suboptimal solutions is also verified. In addition, it is shown that the achievable rate saturates if P_I increases beyond the boundary P_S/l_{21} provided that P_S is fixed. The detailed performance comparison is depicted in the right hand side subfigure for the case of $P_S = 15$ dB. It is verified that the proposed optimal and suboptimal strategies outperform the reference ones especially for small values of P_I.

Figure 4.6 examines the effect of the primary transmit power P_1 on the achievable secondary rate R. We set in left hand side subfigure $P_S = 10$ dB and $P_I = 0$ dB. We compare the performance of all proposed strategies and the reference ones. The near-optimality of the proposed strategies is validated over a large region of P_1. Compared to our strategies, the reference algorithm 1 suffers from a performance degradation in the high P_1 region due to its ignorance of P_1. We remark that the proposed algorithms require no additional instantaneous CSI compared to the reference algorithm 1. Similar observations can be made in the right hand side subfigure with $P_S = 15$ dB and $P_I = 5$ dB are plotted. Furthermore, both subfigures validate the accuracy of the analytical results for the suboptimal strategies.

4.3 Power Control with Outage Probability Constraint

In this subsection, we investigate the gain achieved by the SUs to exploit the partial CSI related to the primary transmission links. Specifically, revising the system model given in Figure 4.1, we assume the ST and the SR have the knowledge of instantaneous CSI g_{22} and g_{12} and statistical CSI l_{11} and l_{21}. Two constraints are imposed on the system design. Firstly, we limit the probability that the instantaneous rate of the primary transmission is smaller than γ_p (bits/s/Hz) below a threshold $\varepsilon_p \in [0,1]$

$$\Pr\left\{\log_2\left(1+\frac{P_1 g_{11}}{P_2(g_{12},g_{22})g_{21}+\sigma_P^2}\right)<\gamma_p\right\}\leq\varepsilon_p \tag{4.87}$$

Secondly, the power of the ST is non-negative and limited by P_S:

$$0\leq P_2(g_{12},g_{22})\leq P_S,\quad \forall\, g_{12}\geq 0,\, g_{22}\geq 0. \tag{4.88}$$

Consequently, the optimization problem aiming at maximizing the average achievable rate of the secondary link subject to the aforementioned constraints is

$$\begin{aligned}\max_{P_2(g_{12},g_{22})}\ & \mathbb{E}_{g_{12},g_{22}}\left\{\ln\left(1+\frac{P_2(g_{12},g_{22})g_{22}}{P_1 g_{12}+\sigma_S^2}\right)\right\}\\ \text{s.t. }\ & \Pr\left\{\log_2\left(1+\frac{P_1 g_{11}}{P_2(g_{12},g_{22})g_{21}+\sigma_P^2}\right)<\gamma_p\right\}\leq\varepsilon_p\\ & 0\leq P_2(g_{12},g_{22})\leq P_S,\quad \forall\, g_{12}\geq 0,\, g_{22}\geq 0.\end{aligned} \tag{4.89}$$

As shown in (4.89), the objective function is a maximization of a concave function of $P_2(g_{12},g_{22})$ and the constraint (4.88) is an affine function of $P_2(g_{12},g_{22})$. Thus, the convexity of (4.89) depends on the convexity of the constraint (4.87) which is unclear in its current form. Therefore, we reformulate the outage probability as

$$\begin{aligned}&\Pr\left\{\log_2\left(1+\frac{P_1 g_{11}}{P_2(g_{12},g_{22})g_{21}+\sigma_P^2}\right)<\gamma_p\right\}\\ &\overset{(a)}{=}\Pr\left\{g_{11}<\frac{\gamma_{\text{th}}\sigma_P^2}{P_1}+\frac{\gamma_{\text{th}}P_2(g_{12},g_{22})}{P_1}g_{21}\right\}\\ &=\mathbb{E}_{g_{12},g_{22}}\left\{\Pr\left\{g_{11}<\frac{\gamma_{\text{th}}\sigma_P^2}{P_1}+\frac{\gamma_{\text{th}}P_2(g_{12},g_{22})}{P_1}g_{21}\,\middle|\,P_2(g_{12},g_{22})\right\}\right\}\\ &\overset{(b)}{=}\mathbb{E}_{g_{12},g_{22}}\left\{1-e^{-\frac{l_{11}\gamma_{\text{th}}\sigma_P^2}{P_1}}\frac{1}{1+\frac{l_{11}\gamma_{\text{th}}}{l_{21}P_1}P_2(g_{12},g_{22})}\right\}\\ &\overset{(c)}{=}1-e^{-\frac{l_{11}\gamma_{\text{th}}\sigma_P^2}{P_1}}\mathbb{E}_{g_{12},g_{22}}\left\{\frac{1}{1+aP_2(g_{12},g_{22})}\right\}\end{aligned} \tag{4.90}$$

where, in (a), we use $\gamma_{\text{th}} = 2^{\gamma_p} - 1$. In (b), the result in [63, Appendix I] is applied. In (c), the deterministic terms are moved out of the expectation operator and the positive $a = l_{11}\gamma_{\text{th}}/(l_{21}P_1)$ is used for notational simplification.

Inserting (4.90) into (4.87), the outage probability constraint is converted into

$$\mathbb{E}_{g_{12},g_{22}}\left\{\frac{1}{1+aP_2(g_{12},g_{22})}\right\} \geq \varepsilon \tag{4.91}$$

where the parameter ε is

$$\varepsilon = e^{\frac{l_{11}\gamma_{\text{th}}\sigma_P^2}{P_1}}(1-\varepsilon_p), \quad \varepsilon \in [0,\,1]. \tag{4.92}$$

The constraint (4.91) is restricting a convex functional on $P_2(g_{12}, g_{22})$ to a value larger than a certain threshold, which is non-convex. Therefore, (4.89) is a non-convex optimization problem. Using the variable t as (4.7), we represent $P_2(g_{12}, g_{22})$ as $P_2(t)$ and reformulate the optimization problem (4.89) as

$$\begin{aligned} \max_{P_2(t)} \quad & R\left(P_2(t)\right) \\ \text{s.t.} \quad & 0 \leq P_2(t) \leq P_S, \quad \forall\, t > 0 \\ & Q_{\text{c}}\left(P_2(t)\right) \geq \varepsilon \end{aligned} \tag{4.93}$$

where $R(\cdot)$ is given in (4.9) and $Q_{\text{c}}(\cdot)$ is defined as

$$Q_{\text{c}}\left(x(t)\right) = \mathbb{E}_t\left\{\frac{1}{1+ax(t)}\right\} \tag{4.94}$$

Due to the equivalence between (4.89) and (4.93), we focus on finding the optimal solution to the non-convex problem (4.93) for the remainder of the subsection.

4.3.1 Optimal Power Allocation

In general, a non-convex optimization problem is difficult to solve. In this section, we aim at designing the optimal solution to the non-convex problem (4.93). To this end, we first show that strong duality holds in spite of non-convexity and thus the KKT conditions are necessary optimality conditions. Using this property, the KKT solutions are derived. The optimality of each KKT solution is then characterized.

Concerning problem (4.93) and $\varepsilon \in [0, 1]$, it is easy to verify that the solution for $\varepsilon = 1$ is $P_2^\star(t) = 0$. Therefore, the remaining task is to solve (4.93) for $0 \leq \varepsilon < 1$. In Appendix A.8, we apply the theorems in [127, 138] and show that strong duality holds between (4.93) and its dual problem. This motivates us to solve the problem in the Lagrangian dual domain.

Since the optimization problem (4.93) has differentiable objective and constraint functions for which strong duality holds, the optimal solution can be attained by solving the KKT conditions [111]. Assuming the non-negative parameter $v^\star$ is the

inverse of an optimal Lagrangian multiplier $\lambda^\star$ and applying the KKT conditions for functional optimization [85,87] to (4.93), we derive the KKT solutions as follows.

Proposition 4.3.1. *Solving the KKT conditions of (4.93) yields*

$$P_2^\star(t) = P_S, \quad \text{if } t \le T_1(v^\star) \tag{4.95}$$
$$P_2^\star(t) = 0, \quad \text{if } t \ge T_2(v^\star) \tag{4.96}$$
$$P_2^\star(t) = x_1(t, v^\star) \ \vee \ P_2^\star(t) = x_2(t, v^\star), \quad \text{if } t \in \mathcal{F}_t \tag{4.97}$$

where

$$x_1(t, v) = \frac{1 - 2v + \sqrt{1 - 4v + 4avt}}{2av} \tag{4.98}$$
$$x_2(t, v) = \frac{1 - 2v - \sqrt{1 - 4v + 4avt}}{2av} \tag{4.99}$$
$$T_1(v) = \frac{(1 + aP_S)^2}{a} v - P_S \tag{4.100}$$
$$T_2(v) = \frac{v}{a}. \tag{4.101}$$

The feasible regions $\mathcal{F}_t$ to constrain $P_2^\star(t) = x_1(t, v^\star) \in (0, P_S)$ and $P_2^\star(t) = x_2(t, v^\star) \in (0, P_S)$ are analyzed in Appendix A.10. If $0 \le \varepsilon \le 1/(1 + aP_S)$, the solution is $P_2^\star(t) = P_S$, cf. (4.91). Thus the calculation of $v^\star$ is not required. If $1/(1 + aP_S) < \varepsilon < 1$, $v^\star$ needs to be selected to let the constraint $Q_c(P_2^\star(t)) \ge \varepsilon$ be satisfied with equality.

Proof. See Appendix A.9. □

Therefore, we only need to focus on deriving the non-trivial optimal solution for $1/(1 + aP_S) < \varepsilon < 1$. Note that KKT conditions are necessary but not sufficient optimality conditions under such circumstances, i.e., the primal and dual optimum are among the KKT solutions. We thus characterize the optimality of the KKT solutions by checking sufficient conditions [14,58].

Lemma 4.3.1. *Given a non-negative $v^\star$, $P_2^\star(t) = x_1(t, v^\star)$ yields a local minimum and $P_2^\star(t) = x_2(t, v^\star)$ yields a local maximum in the corresponding feasible region of t given in Appendix A.10.*

Proof. See Appendix A.11. □

Proposition 4.3.1 combined with Lemma 4.3.1 implies that the optimal solution to (4.93) is made of three functions: two functions at the boundary points, i.e., $P_2^\star(t) = P_S$ and $P_2^\star(t) = 0$, and the function at the interior points $P_2^\star(t) = x_2(t, v^\star)$. The regions of t for these functions are given in (4.95), (4.96), and Appendix A.10, respectively. We assume that the optimal solution is a piecewise continuous function which is a common assumption in optimal control theory, e.g., [74]. In addition, there is no removable discontinuity since this kind of discontinuity can be removed to make the function continuous at the discontinuity. Therefore, the finite number of discontinuities are located in the corresponding feasible region of t for each function. Figure 4.7

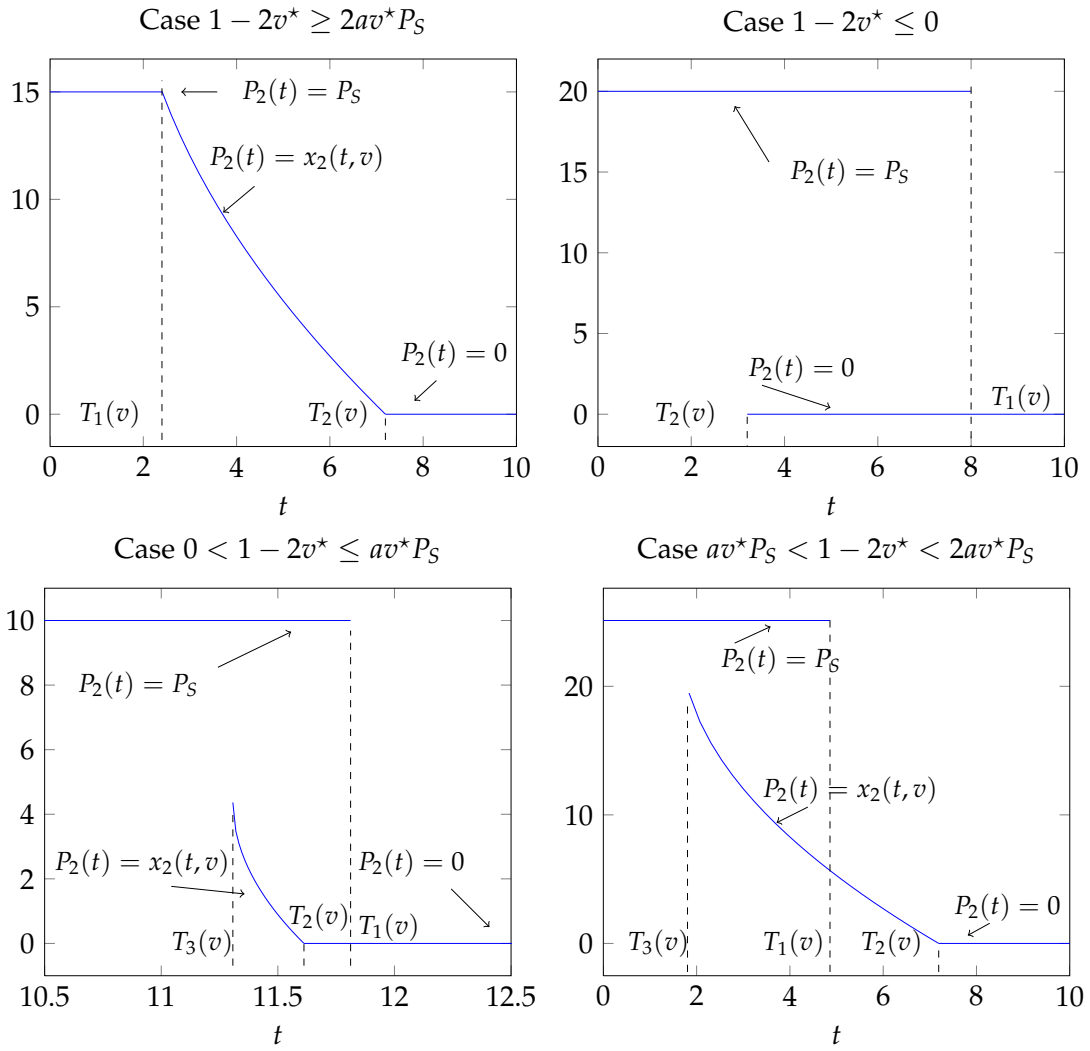

Figure 4.7: Exemplified candidate solutions for different cases of $v^\star$. The functions $T_1(v)$, $T_2(v)$, and $T_3(v)$ are given in (4.100), (4.101), and (A.51), respectively.

exemplifies the candidate solutions for the four cases of $v^\star$. The feasible regions of t for the three functions are not mutually exclusive under the following circumstances:

- The regions of t for $P_2^\star(t)=P_S$ and $P_2^\star(t)=0$ overlap if $1-2v^\star \leq av^\star P_S$.
- The regions of t for $P_2^\star(t)=P_S$ and $P_2^\star(t)=x_2\left(t,v^\star\right)$ overlap if $0<1-2v^\star<2av^\star P_S$.

Due to the fact that the discontinuities can be arbitrarily located in the corresponding feasible region of t for each function and t is continuous, there is an infinite number of candidate solutions. In order to further reduce the set of candidate optimal solutions, we provide an additional characteristic of the optimal solution to (4.93).

Lemma 4.3.2. *The optimal solution $P_2^\star(t)$ to (4.93) is monotonically decreasing in t.*

Proof. See Appendix A.12. □

Lemma A.12 indicates that the optimal power allocation strategy is to allocate a large amount of power when the CSI of the ST-SR link to interference plus noise ratio is high and allocate a small amount of power when this ratio is low. This property is reasonable since the ST needs to allocate more power during favorable channel states in order to maximize the achievable rate of the secondary link while keeping the outage probability of the PU under a certain limit.

Remark: Lemma 4.3.1 can be partly verified by Lemma 4.3.2. More particularly, a solution containing $P_2^\star(t) = x_1(t, v^\star)$ is not the optimal solution since $x_1(t, v^\star)$ is an increasing function of t. However, Lemma 4.3.1 additionally shows that a solution containing $P_2^\star(t) = x_2(t, v^\star)$ might be a optimal solution. This part cannot be validated by Lemma 4.3.2 since it only provides a *necessary* optimality condition tailored to the problem (4.93).

Given the above-mentioned characteristics of the optimal solution, we present the optimal solution of (4.93) in the following theorem:

Theorem 4.3.1. *The optimal solution of (4.93) is given as*
1) $P_2^\star(t) = P_S$, *if* $0 \le \varepsilon \le 1/(1 + aP_S)$.
2) $P_2^\star(t) = 0$, *if* $\varepsilon = 1$.
3) If $1/(1 + aP_S) < \varepsilon < 1$, *the optimal solution* $P_2^\star(t)$ *is divided into three cases:*

- *If* $1 - 2v^\star \ge 2av^\star P_S$,

$$P_2^\star(t) = \begin{cases} P_S, & \text{if } 0 \le t \le T_1(v^\star) \\ x_2(t, v^\star), & \text{if } T_1(v^\star) < t < T_2(v^\star) \\ 0, & \text{otherwise.} \end{cases} \tag{4.102}$$

  The value of $v^\star$ *is chosen to let the equality* $Q_c(P_2^\star(t)) = \varepsilon$ *hold.*

- *If* $1 - 2v^\star \le 0$,

$$P_2^\star(t) = \begin{cases} P_S, & \text{if } 0 \le t \le T_2(v^\star) \\ 0, & \text{otherwise.} \end{cases} \tag{4.103}$$

  The value of $v^\star$ *is chosen to let the equality* $Q_c(P_2^\star(t)) = \varepsilon$ *hold.*

- *If* $0 < 1 - 2v^\star < 2av^\star P_S$, *the solution* $P_2^\star(t)$ *is a function of* $v^\star$ *where*

$$v^\star = \arg\max_v \{R(P_2(t))\} \tag{4.104}$$

  with

$$P_2(t) = \begin{cases} P_S, & \text{if } 0 \le t \le \gamma(v) \\ x_2(t, v), & \text{if } \gamma(v) < t < T_2(v) \\ 0, & \text{otherwise.} \end{cases} \tag{4.105}$$

where the non-negative $\gamma(v)$ is in $[T_3(v), T_1(v)]$ and it is chosen to satisfy $Q_c\left(P_2^\star(t)\right) = \varepsilon$. The functions $T_1(v)$, $T_2(v)$, and $T_3(v)$ are given in (4.100), (4.101), and (A.51), respectively.

Proof. It is easy to verify that the optimal solution is $P_2^\star(t) = P_S$ and $P_2^\star(t) = 0$ for $0 \leq \varepsilon \leq 1/(1 + aP_S)$ and $\varepsilon = 1$, respectively. Therefore, we only focus on the case $1/(1 + aP_S) < \varepsilon < 1$. In the following, the optimal solution is provided as a function of $v^\star$. Incorporating them into the outage probability constraint (4.91) yields the left side as the function of $v^\star$ only. Then $v^\star$ can be obtained by solving (4.91) with equality.

1. $1 - 2v^\star \geq 2av^\star P_S$:

   The case is exemplified in the upper-left subfigure in Figure 4.7. We can verify that $T_1(v^\star) < T_2(v^\star)$ holds. The solutions $P_2^\star(t) = 0$, $P_2^\star(t) = P_S$, and $P_2^\star(t) = x_2(t, v)$ have non-overlapping feasible regions of t. Therefore, the optimal solution is given in (4.102) as the *unique* piecewise continuous function consisting of $P_2^\star(t) = 0$, $P_2^\star(t) = P_S$, and $P_2^\star(t) = x_2(t, v)$.

2. $1 - 2v^\star \leq 0$:

   One example under this case is given in the upper-right subplot in Figure 4.7. For this case $T_2(v^\star) < T_1(v^\star)$ holds. The optimal solution is a piecewise continuous function combined with $P_2^\star(t) = P_S$ and $P_2^\star(t) = 0$ with the feasible region of t as $t \leq T_1\left(v^\star\right)$ and $t \geq T_2\left(v^\star\right)$, respectively. The discontinuity can thus be chosen in $[T_2\left(v^\star\right), T_1\left(v^\star\right)]$. However, if the optimal power allocation strategy is such a binary power switching scheme, there exists a unique discontinuity to let the outage probability constraint be satisfied with equality, i.e., the value of the discontinuity is fixed. Without loss of generality, we give the optimal solution in (4.103) where the discontinuity is located at $T_2\left(v^\star\right)$.

3. $0 < 1 - 2v^\star < 2av^\star P_S$:

   Two different examples are given in the lower-left and lower-right subplots in Figure 4.7 in which $T_2(v^\star) \leq T_1(v^\star)$ and $T_1(v^\star) < T_2(v^\star)$ holds, respectively. The feasible regions of t for $P_2^\star(t) = P_S$, $P_2^\star(t) = 0$, and $P_2^\star(t) = x_2(t, v)$ are $t \leq T_1\left(v^\star\right)$, $t \geq T_2\left(v^\star\right)$, and $T_3\left(v^\star\right) < t < T_2\left(v^\star\right)$, respectively, and they overlap. Therefore, the optimal solution can be a piecewise continuous monotonically decreasing function composed of the three functions in which the discontinuities are located at arbitrary points in the overlapping region. The influence on the objective function of the variation of the discontinuities is difficult to evaluate, resulting in difficulty in further reducing the search space of the optimal solution. We give the solution as a function of v and claim that the optimal solution can be obtained by solving an optimization problem over a single bounded variable v.

   - $T_1\left(v\right) \geq T_2\left(v\right)$: The lower-left subplot in Figure 4.7 indicates one example.
     a) For a discontinuity at $\gamma(v)$ when switching from $P_2^\star(t) = P_S$ to $P_2^\star(t) = x_2(t, v)$ for an increase of t, it is located at an arbitrary point in the region $[T_3\left(v\right), T_2\left(v\right)]$. Note that there is no discontinuity switching

from $P_2^\star(t) = x_2(t,v)$ to $P_2^\star(t) = P_S$ for an increase of t since it violates the optimality condition in Lemma 4.3.2. In addition, the candidate solution switches from $P_2^\star(t) = x_2(t,v)$ to $P_2^\star(t) = 0$ at $T_2(v)$. Due to Lemma 4.3.2, only $P_2^\star(t) = 0$ holds for $t \geq T_2(v)$. The solution is expressed in (4.105) with $\gamma(v) \in [T_3(v), T_2(v)]$ and $\gamma(v)$ is chosen to satisfy the constraint $Q_c(P_2(t)) = \varepsilon$.

b) For a discontinuity at $\gamma(v)$ when switching from $P_2^\star(t) = P_S$ to $P_2^\star(t) = 0$ for an increase of t, then $\gamma(v)$ is located at an arbitrary point in the region $[T_2(v), T_1(v)]$. Similar to the last case, only $P_2^\star(t) = 0$ holds for $t \geq \gamma(v)$ due to the optimality condition in Lemma 4.3.2. The solution is also expressed in the form of (4.105) with $\gamma(v) \in [T_2(v), T_1(v)]$ and $\gamma(v)$ is chosen to let the constraint $Q_c(P_2(t)) = \varepsilon$ be satisfied.

Combined with the above two cases, the power allocation strategy is given in (4.105) with $\gamma(v) \in [T_3(v), T_1(v)]$. The value $\gamma(v)$ should be chosen to satisfy $Q_c(P_2(t)) = \varepsilon$.

- $T_1(v^\star) < T_2(v^\star)$: As shown in the exemplified in the lower-right subplot in Figure 4.7, a discontinuity at $\gamma(v)$ when switching from $P_2^\star(t) = P_S$ to $P_2^\star(t) = x_2(t,v)$ for an increase of t can be located at an arbitrary point in the region $[T_3(v), T_1(v)]$. According to Lemma 4.3.2, there exists no discontinuity when switching from $P_2^\star(t) = x_2(t,v)$ to $P_2^\star(t) = P_S$. Furthermore, the candidate solution switches from $P_2^\star(t) = x_2(t,v)$ to $P_2^\star(t) = 0$ at $T_2(v)$. For $t \geq T_2(v)$, the solution is $P_2^\star(t) = 0$. The solution also matches the form in (4.105). The value $\gamma(v) \in [T_3(v), T_1(v)]$ is chosen to satisfy the constraint $Q_c(P_2(t)) = \varepsilon$.

The proof is concluded by combining the results discussed in each case. □

For the third case $1/(1+aP_S) < \varepsilon < 1$ in Theorem 4.3.1, the value of $v^\star$ is critical to determine the optimal solution. For example, $P_2^\star(t)$ in (4.102) and (4.103) is a function of $v^\star$ since $T_1(v^\star)$, $T_2(v^\star)$, $T_3(v^\star)$, and $x_2(t,v^\star)$ are all functions of $v^\star$, incorporating them into $Q_c(P_2^\star(t))$ results in $Q_c(\cdot)$ being dependent on $v^\star$. We can then obtain $v^\star$ by solving $Q_c(P_2^\star(t)) = \varepsilon$. Similarly, $P_2(t)$ in (4.105) is also provided as a function of v. Taking it into $Q_c(P_2(t))$ and solving $Q_c(P_2(t)) = \varepsilon$ yields a feasible v. Incorporating v into (4.105), we obtain a feasible solution $P_2(t)$. Finally, searching over all feasible v, we select $v^\star$ corresponding to $P_2^\star(t)$ that results in the largest secondary achievable rate in (4.104). Briefly speaking, (4.104) is a nonlinear maximization problem over *a single bounded variable* v for $0 < 1 - 2v < 2avP_S$. Some optimization toolboxes can be applied for its efficient implementation, e.g., *fminbnd* in Matlab©.

4.3.2 Suboptimal Power Allocation

The optimal strategy requires a complex computation for the determination of $v^\star$, especially for the case $0 < 1 - 2v^\star < 2av^\star P_S$ since the optimization (4.104) is required. Furthermore, the achievable performance is hard to evaluate analytically. This necessitates the development of low complex strategies with performance analysis.

4.3.2.1 Suboptimal Power Allocation Strategy I

The suboptimal strategy I results from simplifications of the optimal strategy $P_2^\star(t)$. The simplifications are twofold: Firstly, the nonlinear part $x_2(t,v)$ in $P_2^\star(t)$ is replaced with its first-order Taylor polynomial $\tilde{x}_2(t,v)$; secondly, the time-consuming optimization in (4.104) is replaced by a selection of v among two extreme cases. The resulting solution $\tilde{P}_2^\star(t)$ is given as follows.

Proposition 4.3.2. *Given a certain non-negative $\tilde{v}^\star$, the suboptimal solution I of (4.93) is*

1. $\tilde{P}_2^\star(t) = P_S$, *if* $0 \le \varepsilon \le 1/(1+aP_S)$.
2. $\tilde{P}_2^\star(t) = 0$, *if* $\varepsilon = 1$.
3. *If* $1/(1+aP_S) < \varepsilon < 1$, *we distinguish between three cases:*
   - *If* $1 - 2\tilde{v}^\star \ge 2a\tilde{v}^\star P_S$,

$$\tilde{P}_2^\star(t) = \begin{cases} P_S, & \text{if } 0 \le t \le T_1(\tilde{v}^\star) \\ \tilde{x}_2(t,\tilde{v}^\star), & \text{if } T_1(\tilde{v}^\star) < t < T_2(\tilde{v}^\star) \\ 0, & \text{otherwise.} \end{cases} \tag{4.106}$$

   - *If* $1 - 2\tilde{v}^\star \le 0$,

$$\tilde{P}_2^\star(t) = \begin{cases} P_S, & \text{if } 0 \le t \le T_2(\tilde{v}^\star) \\ 0, & \text{otherwise.} \end{cases} \tag{4.107}$$

   - *If* $0 < 1 - 2\tilde{v}^\star < 2a\tilde{v}^\star P_S$, *the solution* $\tilde{P}_2^\star(t)$ *is chosen among two candidate solutions* $\tilde{P}_{2,1}(t)$ *and* $\tilde{P}_{2,2}(t)$, *that results in the larger objective function*

$$\tilde{P}_2^\star(t) = \arg\max_{\tilde{P}_2(t)} \left\{ R(\tilde{P}_2(t)) : \tilde{P}_2(t) \in \left\{ \tilde{P}_{2,1}(t), \tilde{P}_{2,2}(t) \right\} \right\} \tag{4.108}$$

   The functions $\tilde{P}_{2,1}(t)$ *and* $\tilde{P}_{2,2}(t)$ *are functions of* $\tilde{v}_1$ *and* $\tilde{v}_2$, *respectively*

$$\tilde{P}_{2,1}(t) = \begin{cases} P_S, & \text{if } 0 \le t \le T_1(\tilde{v}_1) \\ \tilde{x}_2(t,\tilde{v}_1), & \text{if } T_1(\tilde{v}_1) < t < T_2(\tilde{v}_1) \\ 0, & \text{otherwise.} \end{cases} \tag{4.109}$$

$$\tilde{P}_{2,2}(t) = \begin{cases} P_S, & \text{if } 0 \le t \le T_3(\tilde{v}_2) \\ \tilde{x}_2(t,\tilde{v}_2), & \text{if } T_3(\tilde{v}_2) < t < T_2(\tilde{v}_2) \\ 0, & \text{otherwise.} \end{cases} \tag{4.110}$$

In (4.106)-(4.110), we use

$$\tilde{x}_2(t,v) = \frac{1}{1-2v}\left(\frac{v}{a} - t\right) \tag{4.111}$$

which is a linear function of t and the functions $T_1(v)$, $T_2(v)$, and $T_3(v)$ are given in (4.100), (4.101), and (A.51), respectively. $\tilde{v}_1$, $\tilde{v}_2$, and $\tilde{v}^\star$ are chosen to satisfy $Q_c\left(\tilde{P}_{2,1}(t)\right) = \varepsilon$, $Q_c\left(\tilde{P}_{2,2}(t)\right) = \varepsilon$ and $Q_c\left(\tilde{P}_2^\star(t)\right) = \varepsilon$, respectively.

Proof. The suboptimal strategy I is derived based on two simplifications of the optimal strategy. Note that the new notation of the Lagrangian variable $\tilde{v}$ is used instead of v in order to indicate a different value of the Lagrangian variable in the strategy. Specifically, we have

1. Linearization of the nonlinear function $\tilde{x}_2(t, v)$:

   Given certain v, we expand $F(t) = \sqrt{1 - 4v + 4avt}$ using the first order Taylor expansion at v/a as

   $$F(t) \approx F\left(\frac{v}{a}\right) + F'\left(\frac{v}{a}\right)\left(t - \frac{v}{a}\right) = 1 - 2v + \frac{2av}{1-2v}\left(t - \frac{v}{a}\right). \tag{4.112}$$

   where $1 - 2v > 0$ holds in order to guarantee the positivity of $x_2(t, v)$, cf. Appendix A.10. Applying (4.112) in the expression of $x_2(t, v)$ in (4.99) yields $\tilde{x}_2(t, v)$ in (4.111). Consequently, the term $x_2(t, v)$ in (4.102) and (4.105) is then replaced by $\tilde{x}_2(t, v)$ in (4.106), (4.109), and (4.110).

2. Simplification of the optimization problem (4.104):

   In order to avoid solving the optimization problem (4.104), we select $\tilde{v}^\star$ from two values $\tilde{v}_1$ and $\tilde{v}_2$ corresponding to two extreme choices of the discontinuities. More particularly, recalling that for the power allocation strategy (4.105), the feasible domain of $\gamma(v)$ is $\gamma(v) \in [T_3(v), T_1(v)]$, we only consider the two cases that $\gamma(v)$ is located at the boundary points. If we choose the discontinuity at $T_1(v)$, $\tilde{v}_1$ is chosen to let the constraint $Q_c\left(\tilde{P}_{2,1}(t)\right) = \varepsilon$ be satisfied and the solution $\tilde{P}_{2,1}(t)$ in (4.109) is obtained. If we choose the discontinuity at $T_3(v^\star)$, the solution is given in (4.110) with $\tilde{v}_2$ chosen to satisfy $Q_c\left(\tilde{P}_{2,2}(t)\right) = \varepsilon$.

Integrating these two steps in the optimal strategy yields the form of the suboptimal strategy I. □

In order to determine $\tilde{v}^\star$ in the third case of Proposition 4.3.2, we need to evaluate the influence of $\tilde{v}$ on $Q_c(\tilde{P}_2(t))$ where the relations between $\tilde{P}_2(t)$ and $\tilde{v}$ are given in (4.106), (4.107), (4.109), or (4.110) for the corresponding regions of $\tilde{v}$. In contrast to the numerical computation of $Q_c(P_2(t))$ in the optimal strategy, we exploit the linear structure of the suboptimal strategy I and derive the analytical form of $Q_c(\tilde{P}_2(t))$. For notational simplicity, the metric $Q_c(\tilde{v}) = Q_c(\tilde{P}_2(t))$ is used in the following proposition which explicitly shows its dependence on $\tilde{v}$.

Proposition 4.3.3. *The function $Q_c(\tilde{v})$ resulting from the suboptimal solution $\tilde{P}_2^\star$ is as follows*

- *If $1 - 2\tilde{v} \geq 2a\tilde{v}P_S$:*

$$\begin{aligned} Q_c(\tilde{v}) = \frac{1}{1 + aP_S} Q_1\left(T_1(\tilde{v})\right) + \frac{1 - 2\tilde{v}}{a}\left(Q_2\left(\zeta, T_2(\tilde{v})\right) - Q_2\left(\zeta, T_1(\tilde{v})\right)\right) \\ + \left(1 - Q_1\left(T_2(\tilde{v})\right)\right). \end{aligned} \tag{4.113}$$

- *If* $1-2\tilde{v}\leq 0$*:*

$$Q_c(\tilde{v}) = 1 - \frac{aP_S}{1+aP_S} Q_1\left(T_2\left(\tilde{v}\right)\right). \tag{4.114}$$

- *If* $0 < 1-2\tilde{v} < 2a\tilde{v}P_S$*:*

  a) Using the power allocation strategy in (4.109), we have

  - *If* $0 < 1-2\tilde{v} \leq a\tilde{v}P_S$*:* $\tilde{P}_2(t)$ *reduces to a switching between* 0 *and* P_S *at* $T_1(\tilde{v})$

$$Q_c\left(\tilde{v}\right) = 1 - \frac{aP_S}{1+aP_S} Q_1\left(T_1\left(\tilde{v}\right)\right). \tag{4.115}$$

  - *If* $a\tilde{v}P_S < 1-2\tilde{v} < 2a\tilde{v}P_S$*: the power allocation strategy* $\tilde{P}_2(t)$ *in (4.109) has the same form as* $\tilde{P}_2(t)$ *in (4.106), thus* $Q_c(\tilde{v})$ *has the same form as (4.113).*

  b) Using the power allocation strategy in (4.110), we have

$$\begin{aligned} Q_c(\tilde{v}) = {} & \frac{1}{1+aP_S} Q_1\left(T_3\left(\tilde{v}\right)\right) + \frac{1-2\tilde{v}}{a}\left(Q_2\left(\zeta, T_2\left(\tilde{v}\right)\right) - Q_2\left(\zeta, T_3\left(\tilde{v}\right)\right)\right) \\ & + \left(1 - Q_1\left(T_2\left(\tilde{v}\right)\right)\right). \end{aligned} \tag{4.116}$$

In (4.113)-(4.116), we use the variable $\zeta = (1-\tilde{v})/a$*. The functions* $Q_1(T)$ *and* $Q_2(\lambda, T)$ *are defined in (A.72) and (A.73).*

Proof. See Appendix A.13. □

Given the power allocation strategy in Proposition 4.3.2, we derive a closed-form expression of the achievable performance. The results are summarized as follows.

Proposition 4.3.4. *The achievable rate* $R\left(\tilde{P}_2^{\star}(t)\right)$ *of the suboptimal strategy I is*

1. $R\left(\tilde{P}_2^{\star}(t)\right) = \lim_{T\to\infty} r\left(T, 1, P_S\right)$*, if* $0 \leq \varepsilon \leq 1/(1+aP_S)$*. This limit can be straightforwardly obtained from (4.24).*

2. $R\left(\tilde{P}_2^{\star}(t)\right) = 0$*, if* $\varepsilon = 1$*, which follows directly with* $\tilde{P}_2^{\star}(t) = 0$*,* $\forall t$*.*

3. *If* $1/(1+aP_S) \leq \varepsilon \leq 1$*, we distinguish among three cases.*

   - $1-2\tilde{v}^{\star} \geq 2a\tilde{v}^{\star}P_S$*:*

$$R\left(\tilde{P}_2^{\star}(t)\right) = r\left(T_1\left(\tilde{v}^{\star}\right), 1, P_S\right) + r\left(T_2\left(\tilde{v}^{\star}\right), \eta, \theta\right) - r\left(T_1\left(\tilde{v}^{\star}\right), \eta, \theta\right). \tag{4.117}$$

   - $1-2\tilde{v}^{\star} \leq 0$*:*

$$R\left(\tilde{P}_2^{\star}(t)\right) = r\left(T_2\left(\tilde{v}^{\star}\right), 1, P_S\right). \tag{4.118}$$

   - $0 < 1-2\tilde{v}^{\star} < 2a\tilde{v}^{\star}P_S$*:*

     a) If the selection problem (4.104) results in $\tilde{P}_2^{\star}(t) = \tilde{P}_{2,1}(t)$ *and* $\tilde{v}^{\star} = \tilde{v}_1$*, then we have*

- *If* $0 < 1 - 2\tilde{v}^\star \leq a\tilde{v}^\star P_S$, *then*

$$R\left(\tilde{P}_2^\star(t)\right) = r\left(T_1\left(\tilde{v}^\star\right), 1, P_S\right). \tag{4.119}$$

- *If* $a\tilde{v}^\star P_S < 1 - 2\tilde{v}^\star < 2a\tilde{v}^\star P_S$, *then* $R\left(\tilde{P}_2^\star(t)\right)$ *has the same form as (4.117).*

b) If the selection problem (4.104) yields $P_2^\star(t) = \tilde{P}_{2,2}(t)$ *and* $\tilde{v}^\star = \tilde{v}_2$, *then we have*

$$R\left(\tilde{P}_2^\star(t)\right) = r\left(T_3\left(\tilde{v}^\star\right), 1, P_S\right) + r\left(T_2\left(\tilde{v}^\star\right), \eta, \theta\right) - r\left(T_3\left(\tilde{v}^\star\right), \eta, \theta\right) \tag{4.120}$$

In above, we use the variables

$$\eta = -\frac{2\tilde{v}^\star}{1 - 2\tilde{v}^\star}, \quad \theta = \frac{\tilde{v}^\star}{\left(1 - 2\tilde{v}^\star\right) a}. \tag{4.121}$$

Proof. See Appendix A.14. □

4.3.2.2 Suboptimal Power Allocation Strategy II

In general, convex optimization problems provide favorable properties in the design of the optimal strategy [111]. Motivated by this fact, we approximate the primal non-convex problem by a convex relaxation and develop the suboptimal strategy II based on the newly-formed convex problem. Specifically, applying Jensen's inequality to $Q_{\mathrm{c}}\left(P_2(t)\right)$, we obtain its lower bound as

$$Q_{\mathrm{c}}\left(P_2(t)\right) \geq \frac{1}{1 + a\mathbb{E}_t\{P_2(t)\}}. \tag{4.122}$$

Consequently, the non-convex constraint (4.91) is approximated by a convex constraint

$$\mathbb{E}_t\{P_2(t)\} \leq \frac{1}{a}\left(\frac{1}{\varepsilon} - 1\right). \tag{4.123}$$

Replacing the second constraint in (4.93) with (4.123) yields

$$\begin{aligned} \max_{P_2(t)} \ & \mathbb{E}_t\left\{\ln\left(1 + \frac{P_2(t)}{t}\right)\right\} \\ \text{s.t.} \ & 0 \leq P_2(t) \leq P_S, \forall t \geq 0 \\ & (4.123). \end{aligned} \tag{4.124}$$

which is a convex optimization problem. Compared to (4.93), the problem (4.124) aims at satisfying a stricter constraint than the primal constraint, yielding a more conservative power allocation strategy. Considering that the constraint (4.123) is in

the form of an average IT constraint, the optimization problem (4.124) is thus solved by the algorithm in Section 4.2.1. The solution is given as

$$\bar{P}_2^\star(t) = \begin{cases} P_S, & \text{if } 0 \le t < \bar{v}^\star - P_S \\ 0, & \text{if } t > \bar{v}^\star \\ \bar{v}^\star - t, & \text{otherwise} \end{cases} \tag{4.125}$$

where $\bar{v}^\star$ is chosen to satisfy the constraint (4.123).

The achievable performance of (4.125) is given as

$$R\left(\bar{P}_2^\star(t)\right) = r\left(\bar{v}^\star - P_S, 1, P_S\right) + r\left(\bar{v}^\star, 0, \bar{v}^\star\right) - r\left(\bar{v}^\star - P_S, 0, \bar{v}^\star\right) \tag{4.126}$$

where $r(T, \alpha, \beta)$ is given in (4.24).

The simulation parameters are given as follows unless explicitly stated. Assuming Rayleigh fading channels for all links, the channel power gain of each fading link g_{ij}, $i, j = 1, 2$, follows an exponential distribution with $l_{11} = 1$, $l_{21} = 2$, $l_{12} = 3$ and $l_{22} = 1$. The noise variances of the primary and secondary links are $\sigma_P^2 = 1$ and $\sigma_S^2 = 1$, respectively. The primary transmit power is $P_1 = 10\,\text{dB}$. In the following, we evaluate the achievable rate in the unit bits/s/Hz.

The performance of four power allocation strategies is compared:

- The optimal strategy given in Theorem 4.3.1.
- The suboptimal strategy I given in Proposition 4.3.2.
- The suboptimal strategy II given in (4.125).
- Non-adaptive power transmission such that the power value is chosen to satisfy all constraints in (4.93) at every channel realization [93].

$$P_{\text{c}}(t) = \min\left\{P_S, \frac{1}{a}\left(\frac{1}{\varepsilon} - 1\right)\right\}. \tag{4.127}$$

Figure 4.8 depicts the achievable secondary rate R versus the peak transmit power limit P_S with different values of the desired outage rate γ_p. The pre-defined outage probability limit is $\varepsilon_p = 20\%$. We have the following observations. Firstly, all proposed strategies outperform the non-adaptive power transmission strategy beyond some value of P_S. The reason is that for a small value of P_S, all four strategies result in the same optimal solution $P_2^\star(t) = P_S$, $\forall t$, when the outage constraint is satisfied, i.e., yielding the same achievable rate. For a large value of P_S, non-adaptive power transmission aims at satisfying all constraints for each channel realization, while others aim at exploiting the channel variations to adapt the power. Secondly, the suboptimal strategy I is near-optimal and outperforms the second suboptimal strategy beyond some value of P_S. The performance gain reflects the advantage of exploiting the primary CSI over the conventional IT constraint. The reason is as follows. If equality in (4.122) holds, then both the primal constraint (4.91) and the approximated constraint

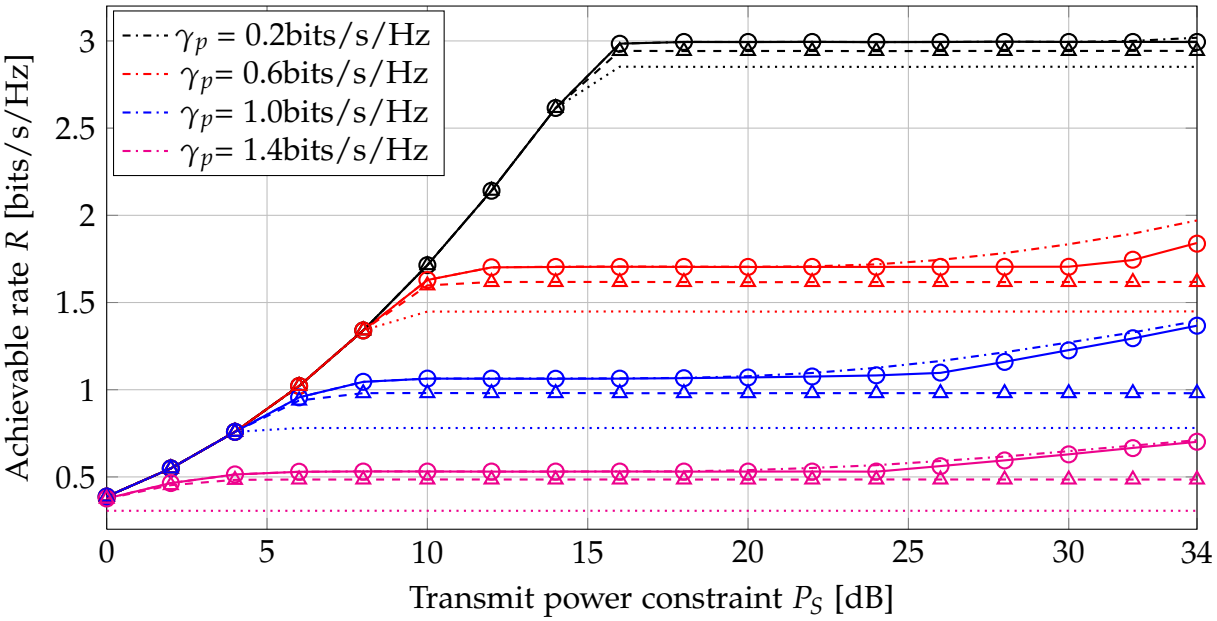

Figure 4.8: Achievable rate R versus P_S for different primary outage rate values γ_p with the outage probability limit is $\varepsilon_p = 20\%$. The lines/markers of the same color share the same setup parameter γ_p. Dash-dotted lines: optimal solution (sim.); solid lines: suboptimal solution I (sim.); circles: suboptimal solution I (analytical); dashed lines: suboptimal solution II (sim.); triangles: suboptimal solution II (analytical); dotted lines: non-adaptive power transmission (sim.).

(4.123) are equivalent, resulting in no performance loss of the suboptimal strategy II. This case happens when $P_2^\star(t) = P_S$, $\forall t$, for a small value of P_S. For a large value of P_S, $P_2^\star(t)$ depends on the channel variation t, i.e., (4.122) is satisfied with strict inequality. Consequently, there exists a performance loss of the suboptimal strategy II. Thirdly, the theoretic performance analysis matches the numerical results for the two suboptimal strategies. Finally, we note that with an increase of P_S, the performance of the optimal and suboptimal strategy I increases first, stays constantly later, and then increases afterwards. This observation is different from that of the suboptimal strategy II in which the performance saturates beyond a certain value of P_S. The reason is roughly explained as follows. For the optimal solution in Theorem 4.3.1, $v^\star$ decreases with an increase of P_S given a fixed γ_p and ε_p in the outage constraint. Thus, $T_1(v^\star)$ in (4.100) can be negative for some values of P_S yielding the condition to let $P_2^\star(t) = P_S$ never satisfied, cf. (4.95). Under such cases, the performance remains constant with an increase of P_S since the peak power constraint is not active. However, since $T_1(v^\star)$ is a quadratic form of P_S and a linear function of $v^\star$, it is possible that an further increase of P_S yields a positive $T_1(v^\star)$ and consequently an active peak power constraint. Hence, the achievable rate increases with an increase of P_S. Similar explanation is applied to the suboptimal solution I. For the suboptimal solution II, $\bar{v}^\star$ in (4.125) decreases with an increase of P_S. Once $\bar{v}^\star$ is smaller than P_S, increasing P_S fur-

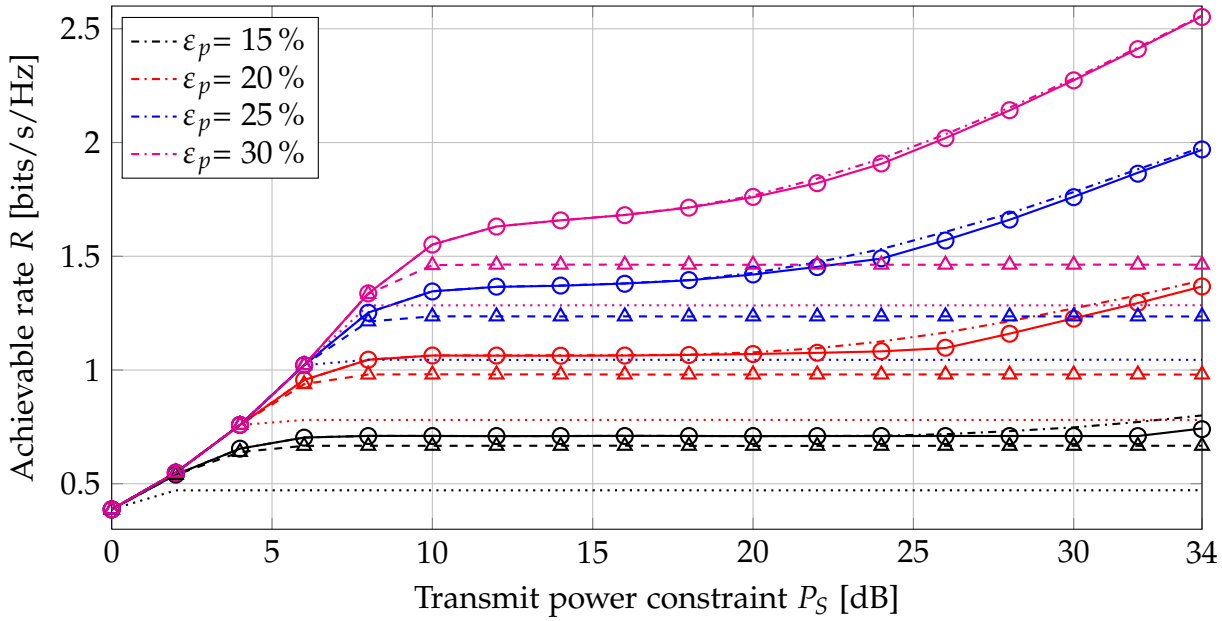

Figure 4.9: Achievable rate R versus P_S for different outage probability limits ε_p; the outage primary rate is $\gamma_p = 1$bits/s/Hz. The explanations of the lines and markers are the same as in Figure 4.8. The lines/markers of the same color share the same setup parameter ε_p.

ther leads the condition of $\bar{P}^\star(t) = P_S$ always be satisfied, cf. (4.125). In other words, its performance is invariant to an increase of P_S beyond some limit. The achievable rate R is plotted versus P_S under different outage probability thresholds ε_p in Figure 4.9. The desired outage capacity is $\gamma_p = 1$ bits/s/Hz. We observe similar trends as in Figure 4.8 for different outage probability requirements.

In order to further visualize reasons of the performance gain of the optimal and the first suboptimal strategies over the second one for a large value of P_S, Figure 4.10 illustrates the resulting outage probability of proposed strategies with different outage probability limits ε_p. The outage capacity is set to $\gamma_p = 1$ bits/s/Hz as in Figure 4.9. For a small value of P_S, all proposed strategies yield the same solution $P_2^\star(t) = P_S$, $\forall t$, thus the same outage probabilities. For a large value of P_S, the suboptimal strategy II yields a more conservative method to protect the PU, i.e., a lower outage probability than ε_p. On the contrary, the outage probability caused by other proposed strategies is shown to meet the requirement. Combining the results shown in Figure 4.9, we conclude that a performance gain w.r.t. secondary achievable rate is achieved at the expense of a larger primary outage probability of the the optimal and the first suboptimal method.

4.4 Summary

The goal in the design of spectrum sharing systems is to reduce the performance degradation to the primary transmission by adjusting the secondary transmission.

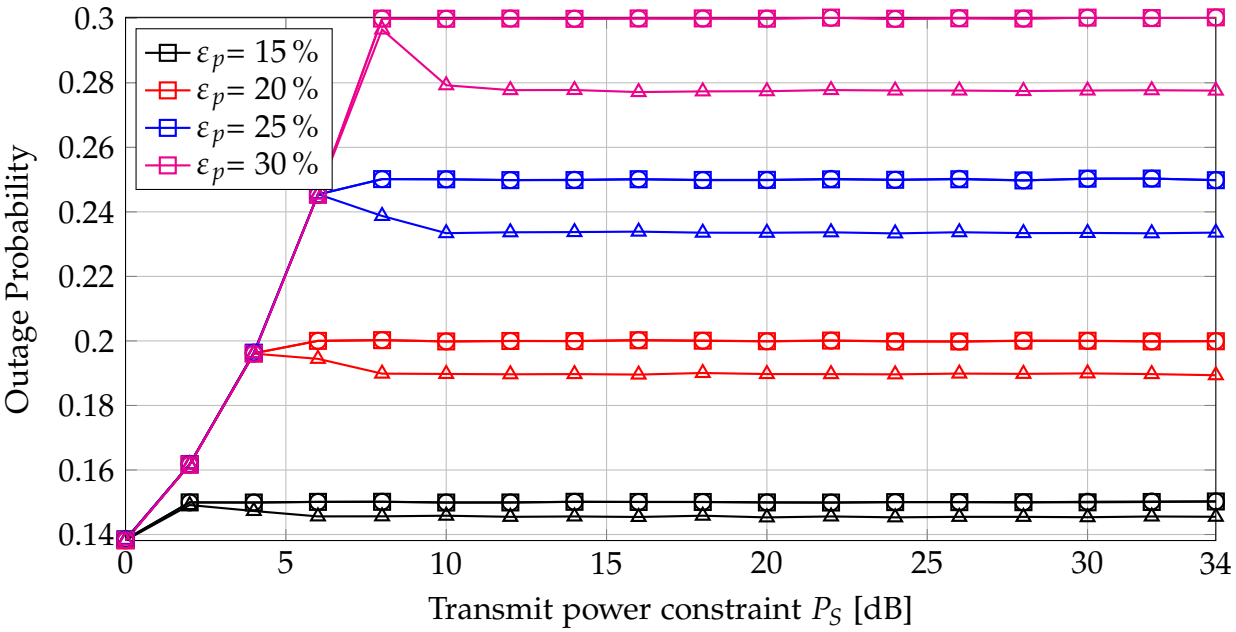

Figure 4.10: Resulting outage probability versus P_S for different ε_p; the outage primary rate is $\gamma_p = 1$bits/s/Hz. The lines/markers of the same color share the same setup parameter ε_p. lines with squares: optimal solution; lines with circles: suboptimal solution I; lines with triangles: suboptimal solution II.

Due to limited cooperation between the primary and secondary systems, the SUs may only have partial CSI related to the primary system in reality. Hence, the design of the optimal transmission strategy and the derivation of the achievable performance of the secondary transmission under such a challenge is imperative.

In this chapter, we have addressed this issue by studying the power allocation of a single-antenna ST aiming at maximizing the achievable rate of the secondary link subject to different QoS constraints of the primary link. Due to realistic restrictions, we assumed the ST has only partial CSI related to PR. Both optimal and computationally efficient suboptimal solutions were developed. Furthermore, the performance of the suboptimal strategies was derived in closed form. Through performance evaluations, the near-optimality of the low-complexity suboptimal solutions was validated. Thus, the achievable performance of the system can be approximately assessed. Moreover, we verified the advantage of additionally exploiting statistical CSI of the primary links by applying the outage probability constraint instead of the IT constraint on the primary transmission. The performance gain can also be approximately evaluated by the derived analytical results. Overall, we have contributed to the quantitative assessment of the achievable performance of the SUs at the expense of a performance degradation at the PUs by exploiting different amounts of available partial CSI. Such results not only characterize the benefit in exploring the spectrum sharing paradigm, but also facilitate the decision in choosing power adaptation strategies.

Chapter 5

Robust Transceiver Optimization

The use of multiple antennas at the transceivers allows for the enhancement of the spectral efficiency by exploiting the spatial DoFs [114,131,135]. Specifically, the signal vector is transmitted towards the desired user with the optimized precoders, while MUI and the effect of fading channels are mitigated at the receiver with the optimized equalizers. Recently, this strategy has been applied to cognitive downlink transmission [60,143]. It effectively improves the spectral utilization and limits the interference.

Due to imperfect or outdated channel estimates and the quantization effect in the limited feedback, erroneous CSI is inevitable. Moreover, in cognitive networks, the limited cooperation between the PUs and the SUs brings new challenges to obtain perfect CSI related to the primary link. Since the presence of imperfect CSI severely degrades the achievable performance [96], or even causes the violation of the interference power limit to the PU, robust design becomes indispensable to deal with the channel uncertainties.

As introduced in Chapter 2, two types of CSI error models are commonly considered: the bounded and the stochastic model. Two corresponding principles of robust design are adopted. One is the worst-case approach to guarantee a certain system performance for any channel realization within the CSI uncertainty region [35, 144, 153]. The other is the stochastic approach which guarantees a certain system performance averaged over all channel realizations within the uncertainty region [151].

This chapter considers the infrastructure-based cognitive downlink transmission where one secondary BS and multiple SUs coexist with the PUs. We target at designing the robust transceiver filters considering both CSI error models: the bounded CSI error in Section 5.1 and the stochastic CSI error in Section 5.2. To ease the computational burden at the SUs, the optimization of both precoders and equalizers is performed at the BS and the equalizers are fed forward to each receiver afterwards. Alternatively, the task of the equalizer optimization can also be assigned to each receiver if it has perfect information of the channel coefficient.

The results presented in this chapter have been published in part by the author in [42–44, 50][1].

[1] In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of RWTH Aachen University's products or services. Internal or personal use of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution, please go to `http://www.ieee.org/publications_standards/publications/rights/rights_link.html` to learn how to obtain a License from RightsLink.

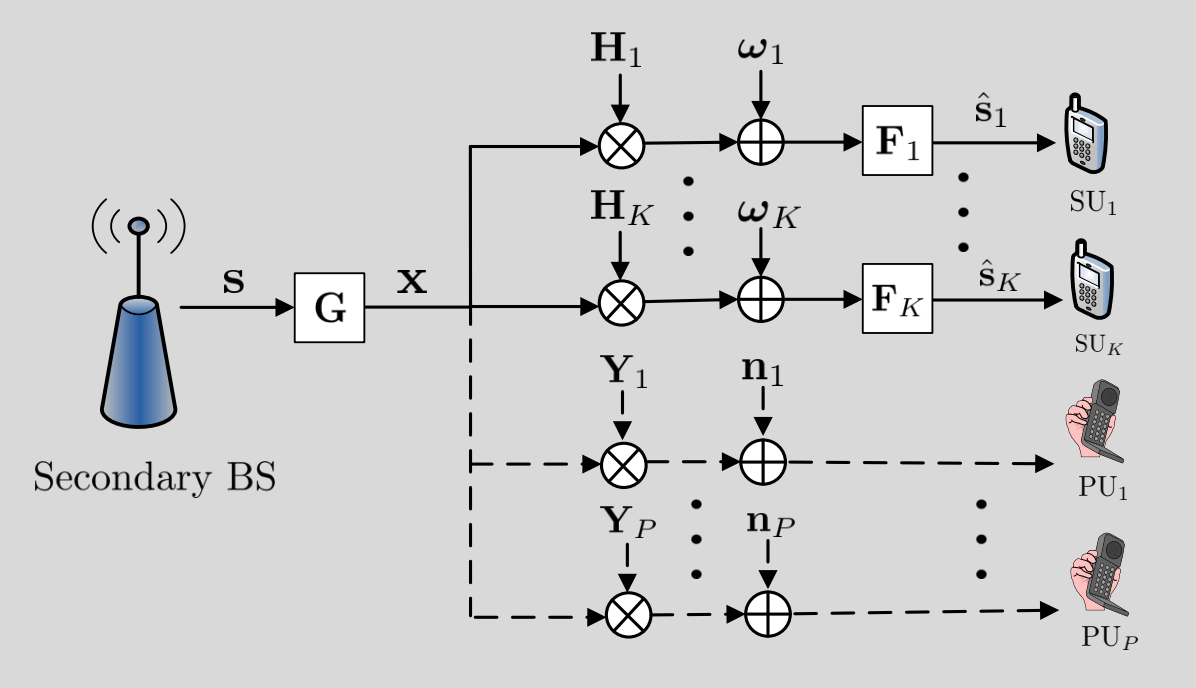

Figure 5.1: Spectrum sharing of the cognitive network with K SUs and P PUs.

5.1 Bounded CSI Error

We consider the downlink CR network depicted in Figure 5.1 where one secondary BS communicating with K SUs. The number of transmit antennas at the BS is given by N_t. Let N_k indicate the number of data streams for the kth user and $\mathbf{s}_k = [s_k^1, \ldots, s_k^{N_k}]^T$ be the symbol vector intended to the kth user. Stacking data vectors into one $N \times 1$ global vector, the data stream for all SUs is $\mathbf{s} = [\mathbf{s}_1^T, \ldots, \mathbf{s}_K^T]^T$ with $N = \sum_{k=1}^{K} N_k$. We assume

$$\mathbb{E}\{\mathbf{s}\mathbf{s}^H\} = \mathbf{I}_N.$$

The precoding matrix for the kth SU is represented by $\mathbf{G}_k \in \mathbb{C}^{N_t \times N_k}$ which yields the precoding matrix at the secondary BS as $\mathbf{G} = [\mathbf{G}_1, \ldots, \mathbf{G}_K]$. The transmit vector at the secondary BS is $\mathbf{x} = \mathbf{G}\mathbf{s}$. Consequently, the average transmitted power at the BS is

$$\mathbb{E}\left[\|\mathbf{x}\|^2\right] = \mathrm{tr}\left\{\mathbf{G}\mathbf{G}^H\right\} = \|\mathbf{G}\|_F^2. \tag{5.1}$$

Assuming the kth SU is equipped with L_k antennas, the downlink flat fading channel from the BS to the kth user is given by $\mathbf{H}_k \in \mathbb{C}^{L_k \times N_t}$. At the receiver side, the linear equalizer $\mathbf{F}_k$ is performed to obtain the recovered data stream $\hat{\mathbf{s}}_k \in \mathbb{C}^{N_k \times 1}$

$$\hat{\mathbf{s}}_k = \mathbf{F}_k\left(\mathbf{H}_k\mathbf{G}\mathbf{s} + \boldsymbol{\omega}_k\right), \quad \forall\, k = 1, \ldots, K \tag{5.2}$$

where $\boldsymbol{\omega}_k \sim \mathcal{CN}(0, \sigma_k^2\mathbf{I})$ indicates the noise at the kth SU with the variance σ_k^2.

There are P PUs coexisting with the cognitive downlink transmission. The pth PU is equipped with L_p antennas, $\forall\ p = 1, \ldots, P$. Denoting the interfering link from the secondary BS to the pth PU as $\mathbf{Y}_p \in \mathbb{C}^{L_p \times N_t}$, the received interference signal at the pth PU from the secondary BS is $\mathbf{z}_p = \mathbf{Y}_p \mathbf{G}\mathbf{s}$ with the interference power given by

$$\mathbb{E}\left[\left\|\mathbf{z}_p\right\|^2\right] = \left\|\mathbf{Y}_p \mathbf{G}\right\|_F^2, \quad \forall\ p = 1, \ldots, P \tag{5.3}$$

We use mean squared error (MSE) as the performance measures. The MSE metric between the symbol vector $\mathbf{s}_k$ and the recovered symbol $\hat{\mathbf{s}}_k$ for the kth user is defined as

$$\mathrm{MSE}_k = \mathbb{E}\left\{\left\|\hat{\mathbf{s}}_k - \mathbf{s}_k\right\|_2^2\right\} = \left\|\mathbf{F}_k \mathbf{H}_k \mathbf{G} - \mathbf{Q}_k\right\|_F^2 + \sigma_k^2 \left\|\mathbf{F}_k\right\|_F^2 \tag{5.4}$$

where

$$\mathbf{Q}_k = \left[\mathbf{0}_{N_k \times \sum_{l=1}^{k-1} N_l}, \mathbf{I}_{N_k}, \mathbf{0}_{N_k \times \sum_{l=k+1}^{K} N_l}\right]. \tag{5.5}$$

The comparison between the MSE measure to other performance measures such as bit error rate (BER) and SINR are elaborated in several works, e.g., [30, 98]. Briefly speaking, the optimization problem based on the BER measure is usually more difficult to handle [98]. Moreover, from the estimation perspective, considering the MSE as a performance measure may be more informative than SINR since it directly reveals the difference between the transmit and recovered signals [30]. Nevertheless, MSE measure is strongly connected to SINR or BER measures [98, 99].

Considering the channel uncertainties, the channel matrix $\mathbf{H}_k$, $\forall\ k = 1, \ldots, K$ is the superposition of two parts

$$\mathbf{H}_k = \hat{\mathbf{H}}_k + \Delta\mathbf{H}_k \tag{5.6}$$

where $\hat{\mathbf{H}}_k$ denotes either the channel estimate or the feedback CSI at the BS and the error matrix of $\mathbf{H}_k$ is given by $\Delta\mathbf{H}_k$. The similar definition is applicable for the channel matrix $\mathbf{Y}_p$

$$\mathbf{Y}_p = \hat{\mathbf{Y}}_p + \Delta\mathbf{Y}_p \tag{5.7}$$

with $\hat{\mathbf{Y}}_p$ as the channel estimate or the feedback CSI at the secondary BS and the error matrix of $\mathbf{Y}_p$ is given by $\Delta\mathbf{Y}_p$, $\forall\, p = 1, \ldots, P$. According to the ellipsoidal uncertainty model [153], the imperfect CSI $\hat{\mathbf{H}}_k$ and $\hat{\mathbf{Y}}_p$ are within the region $\mathcal{H}_k$ and $\mathcal{Y}_p$, respectively

$$\mathcal{H}_k = \left\{\mathbf{H}_k | \mathbf{H}_k = \hat{\mathbf{H}}_k + \Delta\mathbf{H}_k, \mathrm{tr}\left(\Delta\mathbf{H}_k \mathbf{C}_k \Delta\mathbf{H}_k^H\right) \leq \varepsilon_k^2\right\} \tag{5.8}$$

$$\mathcal{Y}_p = \left\{\mathbf{Y}_p | \mathbf{Y}_p = \hat{\mathbf{Y}}_p + \Delta\mathbf{Y}_p, \mathrm{tr}\left(\Delta\mathbf{Y}_p \mathbf{D}_p \Delta\mathbf{Y}_p^H\right) \leq \zeta_p^2\right\} \tag{5.9}$$

where *a priori* known matrices $\mathbf{C}_k$ and $\mathbf{D}_p$ are positive definite Hermitian matrices that define the shape of the uncertainty regions. The sizes of the uncertainty region of $\mathbf{H}_k$ and $\mathcal{Y}_p$ are quantized by ε_k and ζ_p, respectively, $k = 1, \ldots, K$ and $p = 1, \ldots, P$.

We aim at minimizing the sum-MSE of all secondary links subject to the transmit power constraint P_0 at the secondary BS and interference power constraint P_p received at the PR. Assuming only erroneous CSI is available at the secondary BS, the precoding matrices $\mathbf{G}$ and equalizer matrices $\mathbf{F}_1, \ldots, \mathbf{F}_K$ need be designed to ensure the performance targets of both the PUs and the SUs is guaranteed for all channel realizations within the uncertainty region through the optimization. The optimization problem is formulated as

$$\begin{aligned} \min_{\mathbf{G},\mathbf{F}_1,\ldots,\mathbf{F}_K} \max_{\mathbf{H}_k \in \mathcal{H}_k} \quad & \sum_{k=1}^{K} \mathrm{MSE}_k \\ \text{s.t.} \quad & \|\mathbf{G}\|_F^2 \le P_0 \\ & \max_{\mathbf{Y}_p \in \mathcal{Y}_p} \ \left\|\mathbf{Y}_p \mathbf{G}\right\|_F^2 \le P_p, \quad p = 1, \ldots, P. \end{aligned} \tag{5.10}$$

Although only the sum-MSE metric is considered here, the algorithm design could be straightforwardly extended to the problems targeted at other MSE-based measures.

The problem (5.10) generalizes the non-convex problem in [131] by extending the CSI uncertainty model from the ball region to the ellipsoidal region and additionally considering the interference constraint encountered in the CR network. Thus, it is also a non-convex optimization problem, and moreover, a semi-infinite problem which is generally intractable [144]. In order to solve this problem, the authors in [131] provided a framework by applying an alternating principle to iteratively optimize the

Algorithm 1 Alternating algorithm for robust optimization with bounded CSI

Initialization: Iteration number index $l \leftarrow 0$, maximum allowable iterations $l_{\max} \leftarrow L_{\max}$, the desired accuracy ξ, the initial receivers $\mathbf{F}_k^{(0)}$ for all $\forall\, k = 1, \ldots, K$.

repeat

$l \leftarrow l + 1$

Update the precoder $\mathbf{G}^{(l)}$ with fixed equalizers $\mathbf{F}_k^{(l-1)}$, $k = 1, \ldots, K$

$$\begin{aligned} \left\{ \widehat{\mathrm{MSE}}_{k\ k=1,\ldots,K}^{(l)},\ \mathbf{G}^{(l)} \right\} \leftarrow \min_{\mathbf{G}} \max_{\mathbf{H}_k \in \mathcal{H}_k} \quad & \sum_{k=1}^{K} \mathrm{MSE}_k^{(l-1)} \\ \text{s.t.:} \quad & \|\mathbf{G}\|_F^2 \le P_0 \\ & \max_{\mathbf{Y}_p \in \mathcal{Y}_p} \ \left\|\mathbf{Y}_p \mathbf{G}\right\|_F^2 \le P_p, \quad p = 1, \ldots, P. \end{aligned} \tag{5.11}$$

Update the equalizers $\mathbf{F}_k^{(l)}$, $k = 1, \ldots, K$ with fixed precoder $\mathbf{G}^{(l)}$:

$$\left\{ \mathrm{MSE}_k^{(l)},\ \mathbf{F}_k^{(l)} \right\} \leftarrow \min_{\mathbf{F}_k} \max_{\mathbf{H}_k \in \mathcal{H}_k} \widehat{\mathrm{MSE}}_k^{(l)} \tag{5.12}$$

until $l \ge L_{\max}$ **OR** the MSE target converges with the precision ξ

precoders and the equalizers. In this section, analogy to [131], we apply the alternating principle to (5.10). Particularly, decomposing (5.10) into two subproblems (5.11) and (5.12) which tackles with the optimization of the precoders and the equalizers, respectively, the iterative algorithm is then proposed in Algorithm 1. In the following, we seek for the efficient algorithm to solve each subproblem. The convergence behavior of Algorithm 1 is also shown based on the convexity of each subproblem. For the sake of brevity, the iteration number index is omitted in the following derivation.

First, we reformulate the subproblems (5.11) and (5.12) using the vectorization property as

$$\mathrm{vec}(\mathbf{ABC}) = \left(\mathbf{C}^T \otimes \mathbf{A}\right) \mathrm{vec}(\mathbf{B}), \tag{5.13}$$

the MSE term of kth user (5.4) can be reformulated to

$$\mathrm{MSE}_k = \|\boldsymbol{\psi}_k + \bar{\boldsymbol{\Psi}}_k \boldsymbol{\delta}_k\|_2^2 + \sigma_k^2 \|\mathbf{F}_k\|_F^2 \tag{5.14}$$

where

$$\boldsymbol{\psi}_k(\mathbf{G}, \mathbf{F}_k) = \mathrm{vec}(\mathbf{F}_k \hat{\mathbf{H}}_k \mathbf{G} - \mathbf{Q}_k) \tag{5.15}$$
$$\bar{\boldsymbol{\Psi}}_k(\mathbf{G}, \mathbf{F}_k) = \mathbf{G}^T \otimes \mathbf{F}_k \tag{5.16}$$
$$\boldsymbol{\delta}_k = \mathrm{vec}(\Delta \mathbf{H}_k). \tag{5.17}$$

Moreover, the interference power at the pth PU is reformulated as

$$\left\|(\hat{\mathbf{Y}}_p + \Delta \mathbf{Y}_p)\mathbf{G}\right\|_F^2 = \left\|\boldsymbol{\phi}_p + \bar{\boldsymbol{\Phi}}_p \boldsymbol{\delta}_p\right\|_2^2 \tag{5.18}$$

with

$$\boldsymbol{\phi}_p = \mathrm{vec}(\hat{\mathbf{Y}}_p \mathbf{G}) \tag{5.19}$$
$$\bar{\boldsymbol{\Phi}}_p = \mathbf{G}^T \otimes \mathbf{I} \tag{5.20}$$
$$\boldsymbol{\delta}_p = \mathrm{vec}(\Delta \mathbf{Y}_p). \tag{5.21}$$

Similarly, the uncertainty regions $\mathcal{H}_k$ and $\mathcal{Y}_p$ in (5.8) and (5.9) are converted to

$$\bar{\mathcal{H}}_k = \left\{\mathbf{H}_k = \hat{\mathbf{H}}_k + \Delta \mathbf{H}_k |\ \left\|\Delta \mathbf{H}_k \tilde{\mathbf{C}}_k\right\|_F = \left\|(\tilde{\mathbf{C}}_k^T \otimes \mathbf{I})\boldsymbol{\delta}_k\right\|_2 \leq \varepsilon_k\right\} \tag{5.22}$$
$$\bar{\mathcal{Y}}_p = \left\{\mathbf{Y}_p = \hat{\mathbf{Y}}_p + \Delta \mathbf{Y}_p |\ \left\|\Delta \mathbf{Y}_p \tilde{\mathbf{D}}_p\right\|_F = \left\|(\tilde{\mathbf{D}}_p^T \otimes \mathbf{I})\boldsymbol{\delta}_p\right\|_2 \leq \zeta_p\right\} \tag{5.23}$$

respectively, where we use the matrix decomposition

$$\mathbf{C}_k = \tilde{\mathbf{C}}_k \tilde{\mathbf{C}}_k^H \tag{5.24}$$
$$\mathbf{D}_p = \tilde{\mathbf{D}}_p \tilde{\mathbf{D}}_p^H \tag{5.25}$$

and the vectorization in (5.13).

Inserting (5.14) and (5.18) into the subproblem (5.11), we have

$$\begin{aligned} \min_{\mathbf{G},\{\tau_k\}_{k=1}^K} \max_{\mathbf{H}_k \in \bar{\mathcal{H}}_k} \quad & \sum_{k=1}^K \tau_k \\ \text{s.t.} \quad & \|\boldsymbol{\psi}_k + \bar{\boldsymbol{\Psi}}_k \boldsymbol{\delta}_k\|_2^2 \le \tau_k \\ & \max_{\mathbf{Y}_p \in \bar{\mathcal{Y}}_p} \left\|\boldsymbol{\phi}_p + \bar{\boldsymbol{\Phi}}_p \boldsymbol{\delta}_p\right\|_2^2 \le P_p, \quad \forall\, p = 1, \ldots, P \\ & \|\mathbf{G}\|_F^2 \le P_0 \\ & \tau_k \ge 0 \quad \forall\, k = 1, \ldots, K \end{aligned} \tag{5.26}$$

where τ_k is an auxiliary slack variable. The term $\sum_{k=1}^K \sigma_k^2 \|\mathbf{F}_k\|_F^2$ is omitted due to the fixed value of $\mathbf{F}_k$, $\forall\, k = 1, \ldots, K$. The subproblem (5.12) is rewritten as

$$\begin{aligned} \min_{\mathbf{F}_k, \{\tau_k\}_{k=1}^K} \max_{\mathbf{H}_k \in \bar{\mathcal{H}}_k} \quad & \tau_k + \sigma_k^2 \|\mathbf{F}_k\|_F^2 \\ \text{s.t.} \quad & \|\boldsymbol{\psi}_k + \bar{\boldsymbol{\Psi}}_k \boldsymbol{\delta}_k\|_2^2 \le \tau_k \\ & \tau_k \ge 0 \quad \forall\, k = 1, \ldots, K \end{aligned} \tag{5.27}$$

The following lemma is useful to reformulate the subproblems into the form of standard convex optimization problem.

Lemma 5.1.1. *([29, Lemma. 2]) Let* $\mathbf{A}$ *be a Hermitian matrix. Then*

$$\mathbf{A} \succeq \mathbf{B}^H \mathbf{D} \mathbf{C} + \mathbf{C}^H \mathbf{D}^H \mathbf{B}, \quad \forall\, \mathbf{D} : \|\mathbf{D}\|_2 \le \varepsilon \tag{5.28}$$

if and only if there exists a $\lambda \ge 0$ *such that*

$$\begin{bmatrix} \mathbf{A} - \lambda \mathbf{C}^H \mathbf{C} & -\varepsilon \mathbf{B}^H \\ -\varepsilon \mathbf{B}^H & \lambda \mathbf{I} \end{bmatrix} \succeq 0. \tag{5.29}$$

Considering the first subproblem (5.26), if we ignore the interference power constraint first, the following result is obtained.

Proposition 5.1.1. *Dropping the interference power constraint, (5.26) is rewritten in the following semi-definite programming (SDP) form as*

$$\begin{aligned} \min_{\mathbf{G},\{\tau_k\}_{k=1}^K,\{\lambda_k\}_{k=1}^K} \quad & \sum_{k=1}^K \tau_k \\ s.t. \quad & \begin{bmatrix} \tau_k - \lambda_k & \boldsymbol{\psi}_k^H & \mathbf{0} \\ \boldsymbol{\psi}_k & \mathbf{I} & -\varepsilon_k \boldsymbol{\Psi}_k \\ \mathbf{0} & -\varepsilon_k \boldsymbol{\Psi}_k^H & \lambda_k \mathbf{I} \end{bmatrix} \succeq 0 \\ & \tau_k \ge 0, \;\; \lambda_k \ge 0, \qquad \forall\, k = 1 \ldots, K \\ & \|\mathbf{G}\|_F \le P_0, \end{aligned} \tag{5.30}$$

with $\mathbf{\Psi}_k = \left(\tilde{\mathbf{C}}_k^{-1}\mathbf{G}\right)^T \otimes \mathbf{F}_k$.

Proof. Ignoring the interference power constraint, the optimization problem (5.26) is reformulated into

$$\begin{aligned} \min_{\mathbf{G},\tau_1,\ldots,\tau_K} \quad & \max_{\mathbf{H}_k \in \bar{\mathcal{H}}_k} \sum_{k=1}^{K} \tau_k \\ \text{s.t.} \quad & \|\boldsymbol{\psi}_k + \bar{\mathbf{\Psi}}_k \boldsymbol{\delta}_k\|_2^2 \leq \tau_k \\ & \|\mathbf{G}\|_F \leq P_0. \\ & \tau_k \geq 0 \quad \forall\, k = 1, \ldots, K. \end{aligned} \tag{5.31}$$

Applying the Schur Complement Theorem [56] to the first constraint in (5.31), it yields

$$\begin{bmatrix} \tau_k & \boldsymbol{\psi}_k^H \\ \boldsymbol{\psi}_k & \mathbf{I}_K \end{bmatrix} + \begin{bmatrix} 0 & \left(\bar{\mathbf{\Psi}}_k \boldsymbol{\delta}_k\right)^H \\ \bar{\mathbf{\Psi}}_k \boldsymbol{\delta}_k & \mathbf{0} \end{bmatrix} \succeq 0. \tag{5.32}$$

We define the matrix

$$\begin{aligned} \mathbf{\Psi}_k &= \bar{\mathbf{\Psi}}_k \left(\tilde{\mathbf{C}}_k^T \otimes \mathbf{I}\right)^{-1} \\ &\overset{(a)}{=} \left(\mathbf{G}^T \otimes \mathbf{F}_k\right)\left(\tilde{\mathbf{C}}_k^{-T} \otimes \mathbf{I}\right) \\ &\overset{(b)}{=} \left(\tilde{\mathbf{C}}_k^{-1}\mathbf{G}\right)^T \otimes \mathbf{F}_k \end{aligned} \tag{5.33}$$

where, in (a), we use the property of the Kronecker product that

$$(\mathbf{A} \otimes \mathbf{B})^{-1} = \mathbf{A}^{-1} \otimes \mathbf{B}^{-1}.$$

In (b), we use

$$(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = (\mathbf{AC} \otimes \mathbf{BD}).$$

Using (5.33), we construct the following matrices

$$\begin{aligned} \mathbf{A} &= \begin{bmatrix} \tau_k & \boldsymbol{\psi}_k^H \\ \boldsymbol{\psi}_k & \mathbf{I} \end{bmatrix} \\ \mathbf{B} &= \begin{bmatrix} \mathbf{0} & \mathbf{\Psi}_k^H \end{bmatrix} \\ \mathbf{C} &= \begin{bmatrix} -1 & \mathbf{0} \end{bmatrix} \\ \mathbf{D} &= \left(\tilde{\mathbf{C}}_k^T \otimes \mathbf{I}\right)\boldsymbol{\delta}_k \\ \varepsilon &= \varepsilon_k. \end{aligned}$$

and then the constraint (5.32) can be written in the form of

$$\mathbf{A} \succeq \mathbf{B}^H\mathbf{DC} + \mathbf{C}^H\mathbf{D}^H\mathbf{B}.$$

According to Lemma 5.1.1, this condition subject to the uncertainty region (5.22) can be reformulated as

$$\begin{bmatrix} \tau_k - \lambda_k & \boldsymbol{\psi}_k^H & \mathbf{0} \\ \boldsymbol{\psi}_k & \mathbf{I} & -\varepsilon_k \boldsymbol{\Psi}_k \\ \mathbf{0} & -\varepsilon_k \boldsymbol{\Psi}_k^H & \lambda_k \mathbf{I} \end{bmatrix} \succeq 0.$$

Consequently, the problem (5.31) is converted to (5.30). □

Analogously, considering the interference power constraint of the pth PU in (5.26), we insert the following matrices in (5.29)

$$\mathbf{A} = \begin{bmatrix} P_p & \boldsymbol{\phi}_p^H \\ \boldsymbol{\phi}_p & \mathbf{I} \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} \mathbf{0} & \boldsymbol{\Phi}_p^H \end{bmatrix}$$
$$\mathbf{C} = \begin{bmatrix} -1 & \mathbf{0} \end{bmatrix}, \qquad \mathbf{D} = \operatorname{vec}\left(\left(\tilde{\mathbf{D}}_p^T \otimes \mathbf{I} \right) \delta_p \right), \quad \varepsilon = \zeta_p.$$

with

$$\boldsymbol{\Phi}_p = \left(\tilde{\mathbf{D}}_p^{-1} \mathbf{G} \right)^T \otimes \mathbf{I} \tag{5.34}$$

and obtain the equivalent form of the interference power constraint within the uncertainty region is

$$\begin{bmatrix} P_p - \eta_p & \boldsymbol{\phi}_p^H & \mathbf{0} \\ \boldsymbol{\phi}_p & \mathbf{I} & -\zeta_p \boldsymbol{\Phi}_p \\ \mathbf{0} & -\zeta_p \boldsymbol{\Phi}_p^H & \eta_p \mathbf{I} \end{bmatrix} \succeq 0, \tag{5.35}$$

with $\eta_p \geq 0, \forall\ p = 1, \ldots, P$.

Combining (5.30) and (5.35), the subproblem (5.26) is represented as

$$\min_{\mathbf{G}, \{\lambda_k\}_{k=1}^K, \{\tau_k\}_{k=1}^K, \{\eta_p\}_{p=1}^P} \quad \sum_{k=1}^K \tau_k \tag{5.36}$$

$$\text{s.t.} \quad \begin{bmatrix} \tau_k - \lambda_k & \boldsymbol{\psi}_k^H & \mathbf{0} \\ \boldsymbol{\psi}_k & \mathbf{I} & -\varepsilon_k \boldsymbol{\Psi}_k \\ \mathbf{0} & -\varepsilon_k \boldsymbol{\Psi}_k^H & \lambda_k I \end{bmatrix} \succeq \mathbf{0}$$

$$\begin{bmatrix} P_p - \eta_p & \boldsymbol{\phi}_p^H & 0 \\ \boldsymbol{\phi}_p & \mathbf{I} & -\zeta_p \boldsymbol{\Phi}_p \\ 0 & -\zeta_p \boldsymbol{\Phi}_p^H & \eta_p \mathbf{I} \end{bmatrix} \succeq \mathbf{0}$$

$$\|\mathbf{G}\|_F^2 \leq P_0, \quad \forall k$$

$$\tau_k \geq 0,\ \lambda_k \geq 0, \quad \forall\ k = 1, \ldots, K$$

$$\eta_p \geq 0, \quad \forall\ p = 1, \ldots, P. \tag{5.37}$$

Next, considering the subproblem (5.12), it is similar to the equalizer optimization in [131] since the additional interference constraint does not affect given the precoding matrix $\mathbf{G}$ fixed. Therefore, the derivation in [131] can be directly applied except the extension from the ball uncertainty model to the ellipsoidal model. The equivalent SDP form of subproblem (5.27) is

$$\begin{aligned} \min_{\mathbf{F}_k, \{\lambda_k\}_{k=1}^K, \{\tau_k\}_{k=1}^K} \quad & \tau_k + \sigma_k^2 \|\mathbf{F}_k\|_F^2 \\ \text{s.t.} \quad & \begin{bmatrix} \tau_k - \lambda_k & \boldsymbol{\psi}_k^H & \mathbf{0} \\ \boldsymbol{\psi}_k & \mathbf{I} & -\varepsilon_k \boldsymbol{\Psi}_k \\ \mathbf{0} & -\varepsilon_k \boldsymbol{\Psi}_k^H & \lambda_k \mathbf{I} \end{bmatrix} \succeq \mathbf{0}. \\ & \lambda_k \geq 0, \ \tau_k \geq 0, \quad \forall\, k = 1, \ldots, K \end{aligned} \tag{5.38}$$

So far, the subproblems in Algorithm 1 is reformulated into the convex SDP form in (5.36) and (5.38) which can be efficiently solved by convex optimization toolboxes [24], [82]. The implementation complexity of the SDP problem is polynomial in the problem size and $\ln(1/\xi)$ where ξ is the required accuracy [88]. Concerning the optimality issue, we show that Algorithm 1 is guaranteed to converge to the local optimum for any initial selection of equalizers. In order to prove it, first we note that (5.11) is feasible for any initialized equalizers. Assuming $\text{MSE}^{(1a)}$ and $\mathbf{G}^{(1)}$ are the sum-MSE and the precoder obtained in the first iteration by solving (5.11), then $\text{MSE}^{(1a)}$ is also one feasible solution for (5.12). Assuming the subsequent solution of (5.12) is given by $\text{MSE}^{(1b)}$, it can only be smaller or equal to $\text{MSE}^{(1a)}$, i.e., $\text{MSE}^{(1b)} \leq \text{MSE}^{(1a)}$. In the second iteration, since $\mathbf{G}^{(1)}$ is always in the feasible set. so the problem (5.11) will find the precoder $\mathbf{G}^{(2)}$ yielding smaller or equal MSE, i.e., $\text{MSE}^{(2a)} \leq \text{MSE}^{(1b)}$. The rest of the iterative process is deduced by analogy. Therefore, Algorithm 1 creates a monotonically non-increasing sequence of the MSE target with the lower-bound equal to zero. This indicates its convergence to the local optimum. However, the global optimum is not guaranteed.

To evaluate the performance, we consider the CR network consisting of one secondary BS with $N_t = 6$ transmit antennas and two SUs. The simulation parameters for two SUs are the same. Each SU is equipped with $L_k = 2$ antennas and transmits $N_k = 2$ data streams. One PU with a single antenna coexists with the secondary downlink transmission. Concerning the CSI error model in (5.22) and (5.23), the shape parameters of the ellipsoidal region $\mathbf{C}_k$ and $\mathbf{D}_p$ are all identity matrices. The error bound parameters ε_k and ζ_p are set to 0.1. All power values in dB are given relative to the reference level of value 1. The transmit power at the secondary BS is bounded by $P_0 = 0$ dB and the interference power received at the PU is constrained by $P_p = -10$ dB. The desired accuracy ξ in simulations is 10^{-4}.

Four transceiver strategies are compare:

- *Perfect CSI*: The strategy is in Algorithm 1 assuming no channel uncertainty.
- *Robust*: The strategy is in Algorithm 1.

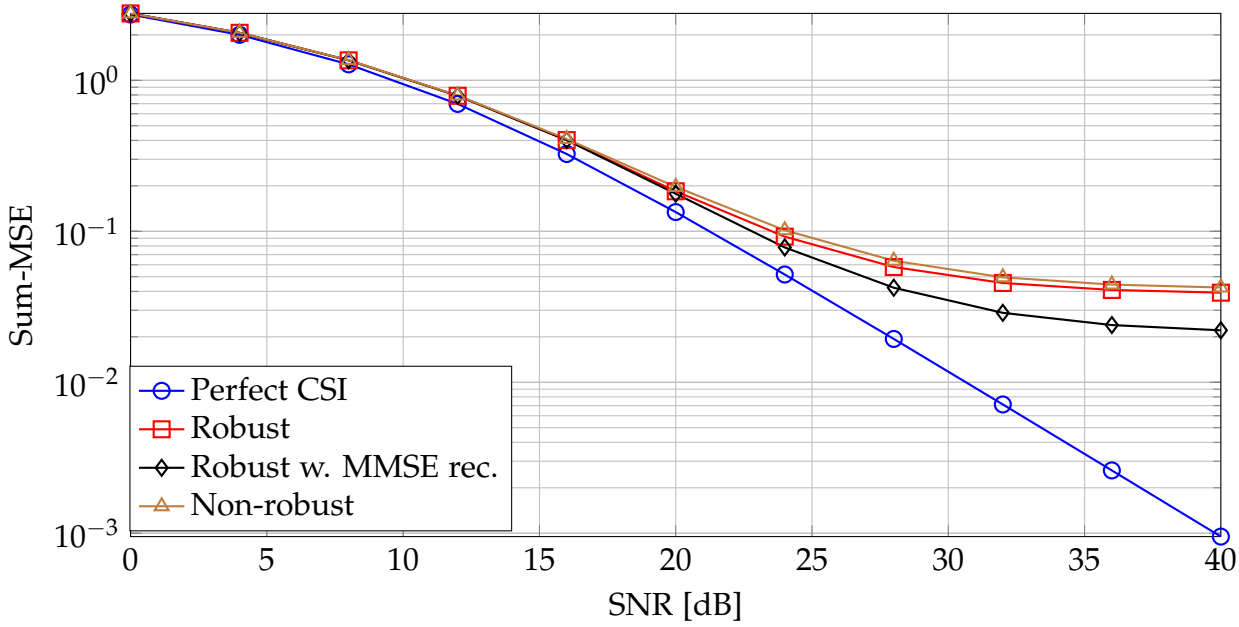

Figure 5.2: Sum-MSE vs. SNR of single CR link.

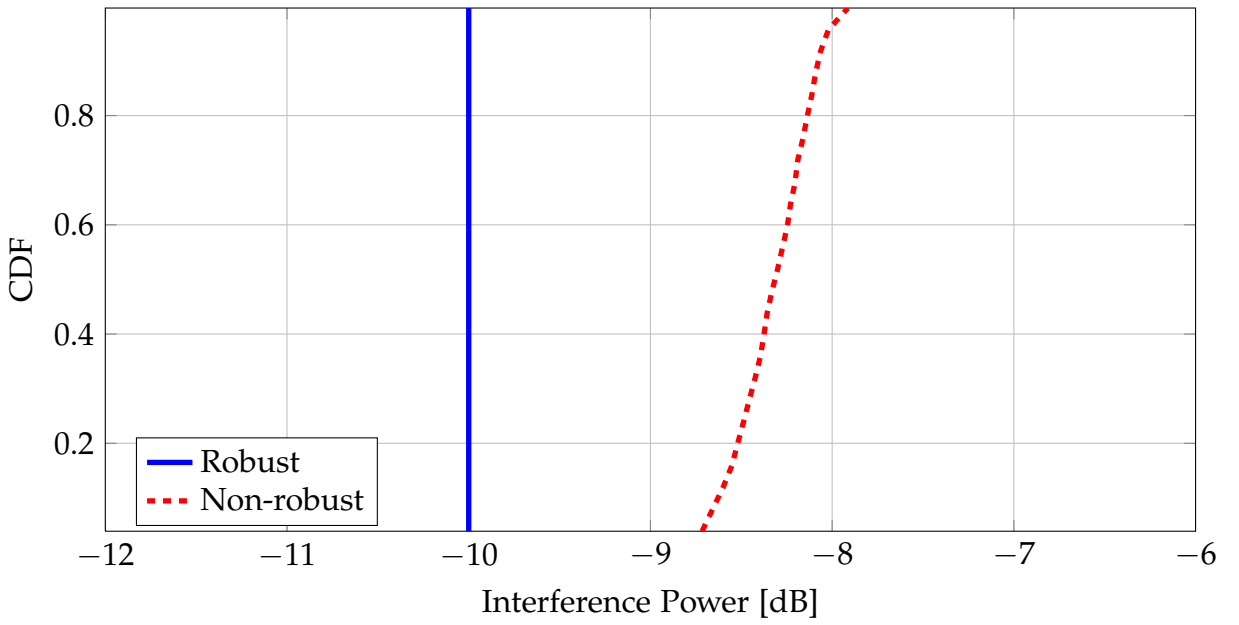

Figure 5.3: CDF of the worst-case interference power at the PR.

- *Robust w. minimum mean squared error (MMSE) rec.*: the strategy is given in Algorithm 1 and the equalizers are calculated according to MMSE criterion afterwards based on the optimal precoder and the worst-case channel coefficients. Similar to [131, Appendix. A], the worst-case channel is obtained which yields the maximum sum-MSE.

- *Non-robust*: the transceiver optimization are calculated by the strategy in Algorithm 1 and neglecting the error in the estimated channel matrices.

In Figure 5.2, we plot the result of sum-MSE versus the SNR of each cognitive link. The proposed robust solution is shown to provide marginal performance gain comparing to the non-robust solution. With the MMSE receiver additionally calculated corresponding to the worst-case channel, the MSE target can be further reduced. We remark that the interference power constraint is often violated in the non-robust case, as shown in Figure 5.3. In the cognitive network, the infringement of the interference power may cause the severe performance degradation to the PU, which should be strictly refrained.

5.2 Stochastic CSI Error

In Section 5.1, we investigate the robust transceiver design with the bounded CSI error and use the worst-case principle. However, it is usually over-conservative from the system perspective. In the following, we address the robust design based on stochastic CSI error and apply the stochastic principle.

We consider a multiple-input multiple-output (MIMO) downlink CR network in Figure 5.1. The input-output relations of the system are identical to that in Section 5.1. The difference is in the CSI error model. Specifically, according to the Kronecker model [68], the channel from the secondary BS to the kth secondary user follows the distribution as

$$\mathrm{vec}(\mathbf{H}_k) \sim \mathcal{CN}\left(\mathbf{0}, \mathbf{R}_k^{Rx} \otimes \mathbf{R}^{Tx}\right), \ \forall\, k = 1, \ldots, K$$

where $\mathbf{R}_{\mathbf{H}_k}^{Rx}$ and $\mathbf{R}^{Tx}$ are the receive and transmit correlation coefficient matrices. By performing MMSE channel estimation [95], the channel is expressed as

$$\mathbf{H}_k = \hat{\mathbf{H}}_k + \Delta\mathbf{H}_k \tag{5.39}$$

where $\hat{\mathbf{H}}_k$ is the estimated channel matrix. The corresponding channel estimation error matrix $\Delta\mathbf{H}_k$ follows as

$$\mathrm{vec}\left(\Delta\mathbf{H}_k\right) \sim \mathcal{CN}\left(\mathbf{0}, \mathbf{R}_{\mathbf{H}_k}^{e,Rx} \otimes \mathbf{R}_{\mathbf{H}_k}^{e,Tx}\right). \tag{5.40}$$

with $\mathbf{R}_{\mathbf{H}_k}^{e,Rx}$ and $\mathbf{R}_{\mathbf{H}_k}^{e,Tx}$ given as the row and column covariance matrices of $\Delta\mathbf{H}_k$, respectively. In general, they can be expressed as a function of $\mathbf{R}_{\mathbf{H}_k}^{Rx}$ and $\mathbf{R}^{Tx}$ depending on the specific channel estimation method [27,95]. A special case that the covariance ma-

trices $\mathbf{R}_{\mathbf{H}_k}^{Rx}$ and $\mathbf{R}^{Tx}$ are the scaled identity matrices has been studied in the supervised thesis [59].

A similar channel model applies to the channel matrix $\mathbf{Y}_p$ from the secondary BS to the pth PU. The corresponding estimated CSI and channel error matrices are denoted by $\hat{\mathbf{Y}}_p$ and $\Delta\mathbf{Y}_p$, $\forall\, p = 1, \ldots, P$

$$\begin{aligned} &\mathbf{Y}_p = \hat{\mathbf{Y}}_p + \Delta\mathbf{Y}_p \\ &\mathrm{vec}(\Delta\mathbf{Y}_p) \sim \mathcal{CN}\left(\mathbf{0}, \mathbf{R}_{\mathbf{Y}_p}^{e,Rx} \otimes \mathbf{R}_{\mathbf{Y}_p}^{e,Tx}\right) \end{aligned} \tag{5.41}$$

The optimization problem aims at minimizing the sum-MSE of the secondary network subject to the transmit power constraint P_0 and the interference power constraint P_p at the pth PU

$$\begin{aligned} \min_{\mathbf{G},\mathbf{F}_1,\ldots,\mathbf{F}_K} \quad & \sum_{k=1}^{K} \mathbb{E}_{\mathbf{H}_k}\left\{\mathrm{MSE}_k\right\} \\ \text{s.t.} \quad & \|\mathbf{G}\|_F^2 \le P_0 \\ & \mathbb{E}_{\mathbf{Y}_p}\left\{\left\|\mathbf{Y}_p\mathbf{G}\right\|_F^2\right\} \le P_p, \;\; \forall\, p = 1, \ldots, P. \end{aligned} \tag{5.42}$$

In order to reveal the convexity of (5.42), we need to express the objective function and the interference constraint as explicit functions of the optimization variables. Specifically, we integrate the stochastic channel model (5.39) and apply [151, Lemma. 1], to the objective function in (5.42), it yields

$$\begin{aligned} &\mathbb{E}_{\mathbf{H}_k}\left\{\mathrm{MSE}_k\right\} \\ &= \left\|\mathbf{F}_k\hat{\mathbf{H}}_k\mathbf{G} - \mathbf{Q}_k\right\|_F^2 + \sigma_k^2\left\|\mathbf{F}_k\right\|_F^2 \\ &\quad + \mathrm{tr}\left\{\mathrm{tr}\left\{\sum_{k=1}^{K}\mathbf{G}_k\mathbf{G}_k^H\left(\mathbf{R}_{\mathbf{H}_k}^{e,Tx}\right)^T\right\}\left\|\mathbf{F}_k\right\|_F^2\mathbf{R}_{\mathbf{H}_k}^{e,Rx}\right\} \\ &= \mathrm{tr}\Bigg\{\mathbf{I}_{N_k} - \mathbf{F}_k\hat{\mathbf{H}}_k\mathbf{G}_k - \mathbf{G}_k^H\hat{\mathbf{H}}_k^H\mathbf{F}_k^H + \sigma_k^2\mathbf{F}_k\mathbf{F}_k^H \\ &\quad + \sum_{l=1}^{K}\mathbf{F}_k\left(\hat{\mathbf{H}}_k\mathbf{G}_l\mathbf{G}_l^H\hat{\mathbf{H}}_k^H + \mathrm{tr}\left(\mathbf{G}_l\mathbf{G}_l^H\left(\mathbf{R}_{\mathbf{H}_k}^{e,Tx}\right)^T\right)\mathbf{R}_{\mathbf{H}_k}^{e,Rx}\right)\mathbf{F}_k^H\Bigg\}. \end{aligned} \tag{5.43}$$

where in (5.43), we use the definition of $\mathbf{Q}_k$ in (5.5). Analogously, by incorporating (5.41) and applying [151, Lemma. 1], the interference power constraint for the pth PU, $\forall\, p = 1, \ldots, P$, is rewritten to

$$\mathbb{E}_{\mathbf{Y}_p}\left\{\left\|\mathbf{Y}_p\mathbf{G}\right\|_F^2\right\} = \mathrm{tr}\left\{\mathbf{A}_p\mathbf{G}\mathbf{G}^H\right\} \tag{5.44}$$

with

$$\mathbf{A}_p = \hat{\mathbf{Y}}_p^H\hat{\mathbf{Y}}_p + \mathrm{tr}\left\{\mathbf{R}_{\mathbf{Y}_p}^{e,Rx}\right\}\left(\mathbf{R}_{\mathbf{Y}_p}^{e,Tx}\right)^T. \tag{5.45}$$

Algorithm 2 SOCP-based robust design

Initialization: Iteration number index $l \leftarrow 0$, maximum allowable iterations $l_{\max} \leftarrow L_{\max}$, the desired accuracy ξ, initial receivers $\mathbf{F}_k^{(0)}$ for all $\forall\, k = 1, \ldots, K$.
repeat
$l \leftarrow l + 1$
Update the precoder $\mathbf{G}^{(l)}$ with fixed equalizers $\mathbf{F}_k^{(l-1)}$, $\forall\, k = 1, \ldots, K$.

$$\left\{ \sum_{k=1}^{K} \mathbb{E}_{\mathbf{H}_k} \left\{ \overline{\mathrm{MSE}}_k^{(l)} \right\}, \mathbf{G}^{(l)} \right\} \leftarrow \min_{\mathbf{G}^{(l)}} \sum_{k=1}^{K} \mathbb{E}_{\mathbf{H}_k} \left\{ \mathrm{MSE}_k^{(l-1)} \right\} \tag{5.46}$$
$$\text{s.t.} \quad \|\mathbf{G}^{(l)}\|_F^2 \leq P_0,$$
$$\mathbb{E}_{\mathbf{Y}_p} \left\{ \left\| \mathbf{Y}_p \mathbf{G}^{(l)} \right\|_F^2 \right\} \leq P_p, \quad \forall p = 1, \ldots, P.$$

Update the equalizers $\mathbf{F}_k^{(l)}$, $\forall\, k = 1, \ldots, K$, with fixed precoder $\mathbf{G}^{(l)}$

$$\left\{ \mathrm{MSE}_k^{(l)}, \ \mathbf{F}_k^{(l)} \right\} \leftarrow \min_{\mathbf{F}_k^{(l)}} \ \mathbb{E}_{\mathbf{H}_k} \left\{ \overline{\mathrm{MSE}}_k^{(l)} \right\} \tag{5.47}$$

until $l \geq L_{\max}$ **OR** the MSE target converges with the precision ξ.

As shown in (5.43), the objective function of (5.42) is non-convex with the joint optimization of the precoders and equalizers which yielding (5.42) to be a non-convex optimization problem. Generally, the optimal solution of such problem is difficult to obtain. In what follows, we design three efficient algorithms to solve it.

5.2.1 SOCP-based Algorithm

Similar to Section 5.1, we apply the alternating principle to solve the optimization problem (5.42). In particular, we first decompose (5.42) into two subproblems (5.46) and (5.47) to compute the precoders and the equalizers separately, and propose an algorithm to solve two subproblems iteratively. The structure of this strategy is outlined in Algorithm 2 named *second-order cone programming (SOCP)-based algorithm*.

In the current form of (5.46) and (5.47), the convexity of each subproblem is unclear. In the following, we address the reformulation into the convex optimization form. Considering the precoder design (5.46) in which the objective function is expressed in (5.43), the term $\sigma_k^2 \|\mathbf{F}_k\|_F^2$ in (5.43) remains constant due to the fixed equalizers, thus, can be omitted. By introducing the auxiliary variables τ_k and r_k, $\forall\, k = 1, \ldots, K$, (5.46) is rewritten as

$$\min_{\mathbf{G}, \{\tau_k\}_{k=1}^K, \{r_k\}_{k=1}^K} \quad \sum_{k=1}^{K} \left(\tau_k + \mathrm{tr} \left\{ r_k \mathbf{F}_k \mathbf{R}_{\mathbf{H}_k}^{e,Rx} \mathbf{F}_k^H \right\} \right) \tag{5.48}$$
$$\text{s.t.} \quad \|\mathbf{F}_k \hat{\mathbf{H}}_k \mathbf{G} - \mathbf{Q}_k\|_F \leq \sqrt{\tau_k},$$

$$
\begin{aligned}
&\|\mathbf{G}\|_F \leq P_T, \\
&\left\|(\mathbf{R}_{\mathbf{H}_k}^{e,Tx})^{T/2}\mathbf{G}\right\|_F \leq r_k \\
&\tau_k \geq 0, \quad \forall\, k = 1, \ldots, K \\
&r_k \geq 0, \quad \forall\, k = 1, \ldots, K \\
&\left\|\mathbf{A}_p^{1/2}\mathbf{G}\right\|_F \leq P_p, \quad \forall\, p = 1, \ldots, P.
\end{aligned}
$$

For the equalizer design in (5.47), the objective function is convex w.r.t. $\mathbf{F}_k$ and independent from $\mathbf{F}_i$, $\forall\, i = 1, \ldots, K$ and $i \neq k$. Therefore, the equalizer of each SU is optimized independently. The closed-form optimum for the kth user $\mathbf{F}_k^\star$ is obtained by calculating the derivative of (5.43) w.r.t. the conjugate of $\mathbf{F}_k$ and setting it to zero

$$
\mathbf{F}_k^\star = \mathbf{G}_k^H \hat{\mathbf{H}}_k^H \mathbf{B}_k^{-1}, \tag{5.49}
$$

with

$$
\mathbf{B}_k = \hat{\mathbf{H}}_k \mathbf{G}\mathbf{G}^H \hat{\mathbf{H}}_k^H + \sigma_k^2 \mathbf{I} + \mathrm{tr}\left\{\mathbf{G}\mathbf{G}^H \left(\mathbf{R}_k^{e,Tx}\right)^T \mathbf{R}_k^{e,Rx}\right\}. \tag{5.50}
$$

Given the SOCP form in (5.48), the first subproblem is solved by the efficient interior-point algorithm using the optimization toolboxes, e.g., CVX [24]. The second subproblem has the analytical solution. Thus, Algorithm 2 solves the problem by iteratively optimizing the two subproblems until the MSE targets of two successive iterations is smaller than a pre-defined precision requirement ξ. This algorithm converges to the *local optimum*. The main computation effect is consumed in solving SOCP problem which is polynomial in the desired accuracy and the problem size [83], e.g., the number of users and antennas. Note that the complexity of solving the SOCP problem is lower than the SDP problem. For the number of elementary arithmetic operations, its upper bound could be calculated according to [13, Section 6.6.2].

5.2.2 Downlink-Based Dual-Loop Algorithm

The computational complexity of the aforementioned SOCP-based algorithm is still relatively high. Alternatively, the MSE minimization problem with a single sum power constraint was solved with a more efficient constrained gradient projection method [15, Section 3.3] in [57]. Motivated by this, we aim at apply this method to the considered problem and propose a low complexity robust solution named downlink-based dual-loop algorithm (DL-DA) in this subsection.

In order to make use of the efficient constrained-gradient projection method, we reformulate the primal problem (5.42) into a two-loop optimization problem where part of the optimization problem has the similar structure as in [57]. Specially, we rewrite the problem (5.42) as

$$
\min_{\mathbf{G},\mathbf{F}_1,\ldots,\mathbf{F}_K} \quad \sum_{k=1}^{K} \mathbb{E}_{\mathbf{H}_k}\left\{\mathrm{MSE}_k\right\} \tag{5.51}
$$

$$\begin{aligned} \text{s.t.} \quad & \text{tr}\left\{\mathbf{A}_0\mathbf{G}\mathbf{G}^H\right\} \leq P_0 \\ & \text{tr}\left\{\mathbf{A}_p\mathbf{G}\mathbf{G}^H\right\} \leq P_p, \ \forall\, p = 1, \ldots, P \end{aligned}$$

where the equivalent form of interference constraint in (5.44) is used. Moreover, according to the sum-power constraint, it holds with $\mathbf{A}_0 = \mathbf{I}$.

We introduce real and nonnegative auxiliary variables μ_i, $\forall\, i = 0, \ldots, P$, and formalize a new problem $g(\{\mu_i\})$ by combining the multiple power constraints

$$\begin{aligned} g(\{\mu_i\}) : \min_{\mathbf{G}, \mathbf{F}_1, \ldots, \mathbf{F}_K} \ & \sum_{k=1}^{K} \mathbb{E}_{\mathbf{H}_k}\{\text{MSE}_k\} \\ \text{s.t.} \quad & \text{tr}\{\mathbf{A}\mathbf{G}\mathbf{G}^H\} \leq P \end{aligned} \tag{5.52}$$

where

$$\mathbf{A} = \sum_{i=0}^{P} \mu_i \mathbf{A}_i \tag{5.53}$$

$$P = \sum_{i=0}^{P} \mu_i \mathbf{P}_i \tag{5.54}$$

The relation between the solutions of (5.51) and (5.52) is addressed in the following propositions.

Proposition 5.2.1. *The optimal value of (5.52) provides a lower bound on that of (5.51) for arbitrary values of the auxiliary variables μ_i, $i = 1, \ldots, P$.*

Proof. This proof follows similarly to the derivation of [145, Proposition 4]. Specifically, we can easily see that if $\mathbf{G}$ is a feasible solution for (5.51), it is also feasible for (5.52). Hence, the feasible region of (5.51) is a subset of that of (5.52). Consequently, the optimal value of (5.52) is equal or smaller than that of (5.51). □

Proposition 5.2.2. *The optimal solution of (5.51) satisfies the KKT optimality conditions of (5.52) given certain μ_i, $\forall\, i = 0, \ldots, P$.*

Proof. We assume the optimal solution of (5.51) is represented by $\mathbf{G}^\star$ and $\lambda_i^\star$, $\forall\, i = 0, \ldots, P$, where λ_i, $\forall\, i = 0, \ldots, P$ are the Lagrange multipliers w.r.t. different power constraints, respectively. We consider the complementary slackness condition in the KKT conditions for the problem (5.51)

$$\lambda_i \left(\text{tr}\left\{\mathbf{A}_i\mathbf{G}\mathbf{G}^H\right\} - P_i\right) = 0, \quad \forall\, i = 0, \ldots, P \tag{5.55}$$

and the complementary slackness condition of (5.52)

$$\lambda \left(\sum_{i=0}^{P} \mu_i \left(\text{tr}\left\{\mathbf{A}_i\mathbf{G}\mathbf{G}^H\right\} - P_i\right)\right) = 0 \tag{5.56}$$

where λ is the Lagrange multiplier associated with the combined power constraint in (5.56). Similar to [145, Proposition 5], if we choose

$$\mathbf{G} = \mathbf{G}^{\star}, \;\; \lambda = 1, \;\; \mu_i = \lambda_i^{\star}, \;\; \forall\, i = 0, \ldots, P \tag{5.57}$$

both (5.55) and (5.56) are satisfied. Similarly, follow the same derivation, the stationarity, primal feasibility, and dual feasibility conditions for both problem are also satisfied accordingly by choosing (5.57). Hence, the optimal solution of (5.51) satisfies the KKT conditions of (5.52) for certain μ_i, $\forall\, i = 0, \ldots, P$. □

In order to obtain $\mu_i, \; i = 0, \ldots, P$, we formulate the optimization problem as

$$\begin{aligned} \max_{\{\mu_i\}_{i=0}^{P}} \quad & g(\{\mu_i\}) \\ \text{s.t.} \quad & \mu_i \geq 0, \;\; i = 0, \ldots, P \end{aligned} \tag{5.58}$$

Based on the result in Proposition 5.2.1, the optimal value of (5.58) serves as a close lower bound to that of (5.51). Moreover, according to Proposition 5.2.2, if the KKT conditions are sufficient for (5.52), the tightness of such lower bound holds, i.e., the optimal value of (5.51) can be attained by solving (5.58).

To summarize, two-loop optimization problem needs to be structured to obtain the solution of (5.51): the inner loop problem (5.52) and the outer loop problem (5.58) form the complete optimization problem. In the following, we propose algorithms to solve both subproblems.

5.2.2.1 Inner loop optimization

Considering the inner loop optimization in (5.52), it has the similar structure as the sum-MSE minimization problem with a single sum power constraint in the conventional multiuser network [57]. Such problem has been efficiently solved by the gradient projection method. Thus, we first reformulate the considered inner-loop problem into a standard form of the sum-MSE minimization problem with a sum power constraint and apply this method.

Specifically, assuming $\mathbf{L} = \mathbf{A}^{-1/2}$ and auxiliary optimization variables are introduced

$$\widetilde{\mathbf{G}}_k = \mathbf{L}^{-H}\mathbf{G}_k, \quad \widetilde{\mathbf{G}} = \sum_{k=1}^{K} \widetilde{\mathbf{G}}_k. \tag{5.59}$$

The next task it to reduce the optimization variable to the equivalent precoder $\widetilde{\mathbf{G}}$. We incorporate (5.59) and (5.49) into the MSE measure (5.43), then take this newly reformulated objective function into (5.52). It yields

$$g(\{\mu_i\}) : \;\; \min_{\widetilde{\mathbf{G}}} \;\; \sum_{k=1}^{K} \mathrm{tr}\{\mathbf{I}_{N_k} - \widetilde{\mathbf{G}}_k^H \mathbf{L}\hat{\mathbf{H}}_k^H \bar{\mathbf{B}}_k^{-1} \hat{\mathbf{H}}_k \mathbf{L}^H \widetilde{\mathbf{G}}_k\} \tag{5.60}$$

$$\text{s.t.} \quad \text{tr}\{\widetilde{\mathbf{G}}\widetilde{\mathbf{G}}^H\} \leq P$$

with

$$\bar{\mathbf{B}}_k = \hat{\mathbf{H}}_k \mathbf{L}\widetilde{\mathbf{G}}\widetilde{\mathbf{G}}^H\mathbf{L}^H\hat{\mathbf{H}}_k^H + \sigma_k^2\mathbf{I} + \text{tr}\left\{\mathbf{L}^H\widetilde{\mathbf{G}}\widetilde{\mathbf{G}}^H\mathbf{L}\left(\mathbf{R}_{\mathbf{H}_k}^{e,Tx}\right)^T\right\}\mathbf{R}_{\mathbf{H}_k}^{e,Rx}. \tag{5.61}$$

We denote the kth precoder in the lth outer loop and the nth inner loop as $\widetilde{\mathbf{G}}_k^{(l,n)}$, $\forall\, k = 1,\ldots,K$. Inheriting the idea of the scaled constrained projection algorithm in [15, Section 3.3], $\widetilde{\mathbf{G}}_k^{(l,n)}$ is optimized iteratively via

$$\widetilde{\mathbf{G}}_k^{(l,n+1)} = \text{Proj}\left[\widetilde{\mathbf{G}}_k^{(l,n)} - \gamma^{(n)}\beta^{(n)}\Delta\widetilde{\mathbf{G}}_k^{(l,n)}\right] \tag{5.62}$$

with γ as the step size and β is the preconditioning scalar chosen to accelerate the convergence speed. Herein we set

$$\beta = \sqrt{\frac{P}{\sum_{k=1}^{K}\left\|\Delta\widetilde{\mathbf{G}}_k\right\|^2}} \tag{5.63}$$

The update variable $\Delta\widetilde{\mathbf{G}}_k$ is

$$\Delta\widetilde{\mathbf{G}}_k := \nabla_k^* \sum_{k=1}^{K} \mathbb{E}_{\mathbf{H}_k}\left\{\text{MSE}_k\right\} \tag{5.64}$$

where ∇_k^* denotes the generation of the Jacobian matrix w.r.t the conjugate of matrix $\widetilde{\mathbf{G}}_k$. In Appendix A.15, $\Delta\widetilde{\mathbf{G}}_k$ is computed as

$$\begin{aligned}\Delta\widetilde{\mathbf{G}}_k = &-\mathbf{L}\hat{\mathbf{H}}_k^H\bar{\mathbf{B}}_k^{-1}\hat{\mathbf{H}}_k\mathbf{L}^H\widetilde{\mathbf{G}}_k + \sum_{i=1}^{K}\mathbf{L}\hat{\mathbf{H}}_k^H\bar{\mathbf{B}}_i^{-1}\hat{\mathbf{H}}_i\mathbf{L}^H\widetilde{\mathbf{G}}_i\widetilde{\mathbf{G}}_i^H\mathbf{L}\hat{\mathbf{H}}_i^H\bar{\mathbf{B}}_i^{-1}\hat{\mathbf{H}}_k\mathbf{L}^H\widetilde{\mathbf{G}}_k \\ &+\sum_{i=1}^{K}\text{tr}\left\{\widetilde{\mathbf{G}}_i^H\mathbf{L}\hat{\mathbf{H}}_i^H\bar{\mathbf{B}}_i^{-1}\mathbf{R}_{\mathbf{H}_k}^{e,Rx}\bar{\mathbf{B}}_i^{-1}\hat{\mathbf{H}}_i\widetilde{\mathbf{G}}_i\right\}\mathbf{L}\left(\mathbf{R}_{\mathbf{H}_k}^{e,Tx}\right)^T\mathbf{L}^H\widetilde{\mathbf{G}}_k.\end{aligned} \tag{5.65}$$

The orthogonal projection in (5.62) is indicated by proj$[\cdot]$. Applying the definition in [15, Section 3.3.1] here, the projection operator guarantees the updated precoders fulfilling the power constraint

$$\begin{aligned}\text{proj}\left[\widetilde{\mathbf{G}}\right] = \arg\min_{\hat{\mathbf{G}}}\left\|\hat{\mathbf{G}} - \widetilde{\mathbf{G}}\right\|_2, \\ \text{s.t. } \text{tr}\left\{\hat{\mathbf{G}}\hat{\mathbf{G}}^H\right\} = P\end{aligned} \tag{5.66}$$

Note that the inequality power constraint in (5.60) is changed to the equality in (5.66) since it is met with equality at the optimum of (5.60). This can be proved by the contradiction. If the power of proj $\left[\widetilde{\mathbf{G}}\right]$ is equal to $\tilde{P} < P$, then the MSE target could

Algorithm 3 DL-DA robust design

1: **Initialization**: Set iteration number $l = 0$, maximal allowable number of iterations l_{max}, the desired accuracy ξ, the initial precoders $\widetilde{\mathbf{G}}_k^{(0,0)}, \forall\, k = 1, \ldots, K$, and the ellipsoid center $\mu_i^{(0)}, \forall\, i = 0, \ldots, P$.
2: **repeat**
3: $l \leftarrow l + 1$, for the fixed $\left\{\mu_i^{(l-1)}\right\}_{i=0}^{P}$.
4: **repeat**
5: $n \leftarrow n + 1$
6: Update $\widetilde{\mathbf{G}}_k^{(l,n)}$ via (5.62), $\forall\, k = 1, \ldots, K$.
7: **until** MSE target converges.
8: Using ellipsoid method to update $\mu_i^{(l)}, \forall\, p = 0, \ldots, P$.
9: **until** $\{\mu_i\}_{i=0}^{P}$ converges within the precision ξ.

be further minimization by scaling each precoder $\widetilde{\mathbf{G}}_k, \forall\, k = 1, \ldots, K$ with $\sqrt{P/\tilde{P}}$. In Appendix A.16, we obtain the result of the projection operation

$$\text{proj}\left[\widetilde{\mathbf{G}}\right] = \sqrt{\frac{P}{\sum_{i=1}^{K}\left\|\widetilde{\mathbf{G}}_i\right\|_F^2}}\widetilde{\mathbf{G}}.$$

5.2.2.2 Outer Loop Optimization

The outer-loop optimization (5.58) is not necessarily differentiable. Thus, the problem can not be directly solved by the gradient ascent method [111]. However, the sub-gradient method applies. After some mathematical manipulations, the sub-gradient $\{s_i^{(j)}\}$ at point $\{\mu_i^{(j)}\}, \forall\, i = 0, \ldots, P$ for the jth iteration is

$$s_i^{(j)} = \text{tr}\left\{\mathbf{G}_{\text{opt}}^{(j)}\left(\mathbf{G}_{\text{opt}}^{(j)}\right)^H \mathbf{A}_i\right\} - P_i \tag{5.67}$$

where $\mathbf{G}_{\text{opt}}^{(j)}$ is the optimal solution of (5.52) with $\mu_i = \mu_i^{(j)}, \forall\, i = 0, \ldots, P$. To maintain the nonnegativity of auxiliary variables, the update of μ_i is selected as

$$\mu_i{}^{(j+1)} = \max\left(f_{\text{update}}\left(\left\{\mu_i{}^{(j)}\right\}_{i=0}^{P}, s_i{}^{(j)}\right), 0\right) \tag{5.68}$$

where $f_{\text{update}}(\cdot)$ is certain update function of $\mu_i, \forall\, i = 0, \ldots, P$ with the sub-gradient determined by (5.67). Herein, we use the ellipsoid method [111].

In summary, the proposed DL-DA algorithm is outlined in Algorithm 3, which is composed of two-loops. First, given the fixed $\mu_i, \forall\, i = 0, \ldots, P$, the inner loop (5.52) is solved by iteratively updating the dual uplink precoding matrices via (5.62) until

convergence. Second, the outer loop (5.58) is searching for the optimal μ_i via the sub-gradient based method (5.68).

5.2.3 Duality-Based Dual-Loop Algorithm

In the last subsections, the non-convex problem (5.42) was solved by either the alternating method or the dual-loop optimization in the downlink system. In general, the downlink problem is usually non-convex and difficult to handle due to the coupling parameters. Alternatively, the uplink problem usually has some favorable attributes, e.g., less coupling of parameters or hidden convexity [62, 145]. In order to take advantage of these properties, the downlink optimization problem can be transformed into the dual uplink problem using uplink–downlink (UL–DL) duality. Such duality shows the reciprocity relationship between the uplink and downlink problems.

In this subsection, we refine the previous algorithms by solving the problem via the exploitation of the developed UL-DL MSE duality, i.e., the same MSE-based targets are achieved in both the primal downlink and the dual uplink problem. Based on this result, a dual-loop algorithm is proposed for the dual uplink problem. We show that the proposed duality-based robust algorithm has faster convergence rate and lower complexity compared to the downlink-based methods.

5.2.3.1 UL–DL MSE Duality with Imperfect CSI

MSE duality under a *single* sum power constraint was studied either with perfect CSI [57] or imperfect CSI [18]. However, in the cognitive network, both the transmit and interference power constraints need to be considered. Therefore, we need to study the duality with multiple power constraints. Inheriting the idea from [145] where SINR duality was established with multiple power constraints, we establish the MSE duality with multiple power constraints and imperfect CSI, which provides the basis of the proposed algorithm. The MSE duality is established in two steps. First, only a single power constraint and imperfect CSI is considered. Second, the above duality is applied to the system with multiple power constraints.

MSE Duality with a Single Power Constraint We consider the downlink and uplink networks depicted in Figure 5.4. Although the network structure is analogous to [145], here we consider the CSI imperfection in modeling the channel matrices. Considering the downlink optimization problem as

$$\begin{aligned} \min_{\mathbf{G},\mathbf{F}_1,\ldots,\mathbf{F}_K} \quad & \sum_{k=1}^{K} \mathbb{E}_{\mathbf{H}_k}\left\{\mathrm{MSE}_k\right\} \\ \text{s.t.} \quad & \mathrm{tr}\{\mathbf{A}\mathbf{G}\mathbf{G}^H\} \leq \tilde{P} \end{aligned} \tag{5.69}$$

where the kth user's MSE is given in (5.4). The power limit is denoted by $\tilde{P}$. The power constraint is the linear function of the transmit covariance matrix $\mathbf{G}\mathbf{G}^H$ scaled

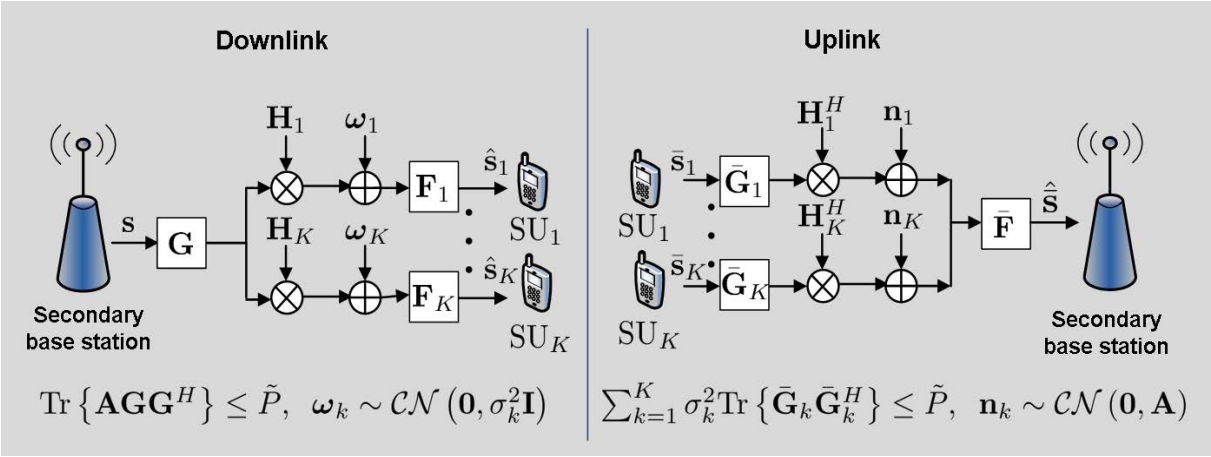

Figure 5.4: Network structure of the primal downlink and the dual uplink with a single power constraint.

with certain matrix $\mathbf{A}$. For example, for the sum power constraint, $\mathbf{A} = \mathbf{I}$ holds; for the interference constraint in (5.44), $\mathbf{A}$ is replaced by $\mathbf{A}_p$ in (5.45).

The dual uplink problem is modeled in the right hand side of Figure 5.4. The precoders and equalizers are exchanged in the downlink network. K SUs transmit signals $\mathbf{s}_1, \ldots, \mathbf{s}_K$ to the secondary BS. The uplink channel is the conjugate transpose of the corresponding downlink channel, i.e., $\mathbf{H}_k^H$ from the kth SU to the BS. The noise covariance matrix for every link is $\mathbf{A}$. The precoders, equalizers and MSE are denoted by $\bar{\mathbf{G}}_k$, $\bar{\mathbf{F}}_k$ and $\overline{\mathrm{MSE}}_k$, respectively. Consequently, the optimization problem is

$$\begin{aligned} \min_{\bar{\mathbf{G}}, \bar{\mathbf{F}}_1, \ldots, \bar{\mathbf{F}}_K} \quad & \sum_{k=1}^K \mathbb{E}_{\mathbf{H}_k} \left\{ \overline{\mathrm{MSE}}_k \right\} \\ \text{s.t.} \quad & \sum_{k=1}^K \sigma_k^2 \mathrm{tr} \left\{ \bar{\mathbf{G}}_k \bar{\mathbf{G}}_k^H \right\} \le \tilde{P} \end{aligned} \tag{5.70}$$

where

$$\begin{aligned} \mathbb{E}_{\mathbf{H}_k} \{ \overline{\mathrm{MSE}}_k \} = \mathrm{tr} \Bigg\{ & \mathbf{I}_{N_k} - \bar{\mathbf{F}}_k \hat{\mathbf{H}}_k^H \bar{\mathbf{G}}_k - \bar{\mathbf{G}}_k^H \hat{\mathbf{H}}_k \bar{\mathbf{F}}_k^H + \bar{\mathbf{F}}_k \mathbf{A} \bar{\mathbf{F}}_k^H \\ & + \sum_{l=1}^K \bar{\mathbf{F}}_k \left(\hat{\mathbf{H}}_l^H \bar{\mathbf{G}}_l \bar{\mathbf{G}}_l^H \hat{\mathbf{H}}_l + \mathrm{tr} \left(\mathbf{R}_{\mathbf{H}_l}^{e,Rx} \bar{\mathbf{G}}_l \bar{\mathbf{G}}_l^H \right) \left(\mathbf{R}_{\mathbf{H}_l}^{e,Tx} \right)^T \right) \bar{\mathbf{F}}_k^H \Bigg\}. \end{aligned} \tag{5.71}$$

The duality of the primal downlink problem (5.69) and the dual uplink problem (5.70) is summarized as follows.

Theorem 5.2.1. *Using the linear relation of the UL-DL precoders and equalizers as*

$$\mathbf{G}_k = \alpha_k \bar{\mathbf{F}}_k^H, \quad \mathbf{F}_k = \alpha_k^{-1} \bar{\mathbf{G}}_k^H, \ \ \forall k = 1, \ldots, K \tag{5.72}$$

where α_k is a positive scaling variable.Both the uplink and downlink systems achieve the same use-wise MSE subject to the same power constraint. Hence, the primal problem (5.69) and the dual uplink problem (5.70) achieve the same MSE region with imperfect CSI by using the same set of the precoders and the equalizers.

Proof. MSE duality indicates the following UL-DL relationship: for any set of $\bar{\mathbf{G}}_k$ and $\bar{\mathbf{F}}_k$ of the dual uplink that achieves certain user-wise MSE, there exists at least one set of $\mathbf{G}_k$ and $\mathbf{F}_k$ of the corresponding downlink system that achieves the same use-wise MSEs under the same sum-power consumption. Considering the linear relation of the downlink and uplink precoders and equalizers in (5.72), in order to keep the same user-wise MSEs in the UL-DL conversion, the conditions $\mathbb{E}_{\mathbf{H}_k}\{\mathrm{MSE}_k\} = \mathbb{E}_{\mathbf{H}_k}\{\overline{\mathrm{MSE}}_k\}$, $\forall\, k = 1, \ldots, K$ should be fulfilled. After some mathematical manipulations, α_k can be solved through the following linear equations:

$$\mathbf{T}\left[\alpha_1^2, \ldots, \alpha_K^2\right]^T = \left[\mathrm{tr}\left\{\sigma_1^2 \bar{\mathbf{G}}_1 \bar{\mathbf{G}}_1^H\right\}, \ldots, \mathrm{tr}\left\{\sigma_K^2 \bar{\mathbf{G}}_K \bar{\mathbf{G}}_K^H\right\}\right]^T \tag{5.73}$$

where $\mathbf{T}$ is shown as follows

$$[\mathbf{T}]_{i,j} = \begin{cases} \mathrm{tr}\left\{\bar{\mathbf{F}}_i \left(\sum\limits_{l=1, l\neq i}^{K} \hat{\mathbf{H}}_l^H \bar{\mathbf{G}}_l \bar{\mathbf{G}}_l^H \hat{\mathbf{H}}_l + \mathrm{tr}\left(\mathbf{R}_{\mathbf{H}_i}^{e,Rx} \bar{\mathbf{G}}_l \bar{\mathbf{G}}_l^H\right)\left(\mathbf{R}_{\mathbf{H}_i}^{e,Tx}\right)^T + \mathbf{A}\right) \bar{\mathbf{F}}_i^H\right\}, & i = j \\ -\mathrm{tr}\left\{\bar{\mathbf{G}}_i^H \left(\hat{\mathbf{H}}_i \bar{\mathbf{F}}_j^H \bar{\mathbf{F}}_j \hat{\mathbf{H}}_i^H + \mathrm{tr}\left\{\bar{\mathbf{F}}_j^H \bar{\mathbf{F}}_j \left(\mathbf{R}_{\mathbf{H}_i}^{e,Tx}\right)^T\right\} \mathbf{R}_{\mathbf{H}_j}^{e,Rx}\right) \bar{\mathbf{G}}_i\right\}, & i \neq j \end{cases}. \tag{5.74}$$

Since $\mathbf{T}$ is the column diagonally dominant matrix, $\mathbf{T}^{-1}$ always exists with all entries nonnegative and the diagonal entries strictly positive. This ensures that α_k has the feasible solution. Adding all equations in (5.73) results in the equivalent power consumption in (5.69) and (5.70),

$$\sum_{l=1}^{K} \sigma_l^2 \mathrm{tr}\left\{\bar{\mathbf{G}}_l^H \bar{\mathbf{G}}_l\right\} = \sum_{l=1}^{K} \mathrm{tr}\{\alpha_l^2 \bar{\mathbf{F}}_l \mathbf{A} \bar{\mathbf{F}}_l^H\} = \mathrm{tr}\{\mathbf{A}\mathbf{G}\mathbf{G}^H\} \tag{5.75}$$

which concludes the proof. □

Aforementioned duality preserves the user-wise MSEs through the UL-DL conversion. Therefore, the duality result is applicable to a group of general optimization problems in which the objective function is an arbitrary linear combination of the user-wise MSEs, e.g., minimizing the worst user-wise MSE averaged over the channel realizations within the uncertainty region.

In the special case of the sum-MSE minimization problems (5.69) and (5.70), only DoF equal to one is required to preserve the same sum-MSE. Thus, α_k is further simplified by setting them to the same value [57]. Computed with the aid of the transmit power constraint (5.75), it gives

$$\alpha_1 = \alpha_2 = \ldots \alpha_K = \sqrt{\frac{\tilde{P}}{\sum_{l=1}^{K} \mathrm{tr}\left(\bar{\mathbf{F}}_l \mathbf{A} \bar{\mathbf{F}}_l^H\right)}}. \tag{5.76}$$

MSE Duality with Multiple Power Constraints The MSE-duality with multiple power constraints is exemplified by considering the problem (5.42). As discussed before, it can be solved through a dual-loop optimization by introducing the auxiliary variables $\mu_i, \forall\, i = 0, \ldots, P$. The inner loop (5.52) is subject to a single power constraint and the outer loop (5.58) aims at the optimization of μ_i, $\forall\, i = 0, \ldots, P$. Applying this idea to the dual uplink problem, MSE duality with multiple power constraints is given as follows: we first combine the multiple power constrains in the downlink system as (5.52) and apply the derived MSE duality to it. Since the problem (5.52) is in the same form as (5.69), the dual uplink problem is shown in (5.70). The second step is to determine the optimal values for nonnegative variables μ_i by solving (5.58), $\forall\, i = 0, \ldots, P$. Finally, the solutions obtained from the dual uplink problem is converted to (5.42) using the linear transformations (5.72). The linear transformation parameters are chosen as the solutions to (5.73). In the special case of sum-MSE minimization problem in (5.42), they are in the form of (5.76).

We remark that, the MSE duality result is the extension of the duality in [57] to the case of imperfect CSI and the duality in [18] to the case of a general form of power constraint as an *arbitrary* linear function of transmit covariance matrices. Applying such duality to the primal optimization problem (5.42) yields the duality-based dual-loop optimization algorithm: the inner-loop in (5.69) and the outer-loop (5.70). In what follows, we propose the efficient method to solve them. Similar to the DL-DA algorithm in Algorithm 3, the dual uplink problem can also be solved by the efficient gradient projection method [15, 57]. The resulting algorithm is named duality-based dual-loop algorithm (DB-DA).

5.2.3.2 Inner-Loop Optimization

We introduce an auxiliary optimization variables $\check{\mathbf{G}}_k$ as

$$\check{\mathbf{G}}_k = \sigma_k \bar{\mathbf{G}}_k. \tag{5.77}$$

The weighted sum of individual power constraint is then rewritten in the form of a standard sum-power constraint. Moreover, given the optimal fixed equalizers, the MSE of the kth user given the optimal precoders is the function of the precoders as

$$\mathbb{E}_{\mathbf{H}_k}\{\overline{\mathrm{MSE}}_k\} = \mathbf{I}_{N_k} - \sigma_k^{-2} \check{\mathbf{G}}_k^H \hat{\mathbf{H}}_k \check{\mathbf{B}}^{-1} \hat{\mathbf{H}}_k^H \check{\mathbf{G}}_k \tag{5.78}$$

where we use $\tilde{\mathbf{A}} = \sum_{i=0}^{P} \mu_i \mathbf{A}_i$ and

$$\check{\mathbf{B}} = \sum_{k=1}^{K} \sigma_k^{-2} \left(\hat{\mathbf{H}}_k^H \check{\mathbf{G}}_k \check{\mathbf{G}}_k^H \hat{\mathbf{H}}_k + \mathrm{tr}\left(\mathbf{R}_{\mathbf{H}_k}^{e,Rx} \check{\mathbf{G}}_k \check{\mathbf{G}}_k^H \right) \left(\mathbf{R}_{\mathbf{H}_k}^{e,Tx} \right)^T \right) + \tilde{\mathbf{A}}.$$

The optimization problem (5.70) is then reformulated as

$$\min_{\{\check{\mathbf{G}}_k\}_{k=1}^K} \quad \sum_{k=1}^{K} \mathbb{E}_{\mathbf{H}_k}\{\overline{\mathrm{MSE}}_k\} \tag{5.79}$$

$$\text{s.t.} \quad \sum_{k=1}^{K} \text{tr}\left(\check{\mathbf{G}}_k \check{\mathbf{G}}_k^H\right) \leq \tilde{P}.$$

Applying the gradient projection method [15] in (5.79), the kth precoder in the lth outer loop and the nth inner loop is updated via

$$\check{\mathbf{G}}_k^{(l,n+1)} = \text{proj}\left[\check{\mathbf{G}}_k^{(l,n)} - \gamma^{(n)}\beta^{(n)}\Delta\check{\mathbf{G}}_k^{(l,n)}\right] \tag{5.80}$$

where γ is the step size. Analogous to the DL-DA, β is the preconditioning scalar chosen as

$$\beta = \sqrt{\frac{\tilde{P}}{\sum_{k=1}^{K}\left\|\Delta\check{\mathbf{G}}_k\right\|^2}} \tag{5.81}$$

for the sake of accelerating the convergence speed [57]. The variable $\Delta\check{\mathbf{G}}_k$ is

$$\begin{aligned}\Delta\check{\mathbf{G}}_k &= \nabla_k^* \sum_{k=1}^{K} \mathbb{E}_{\mathbf{H}_k}\{\overline{\text{MSE}}_k\} \\ &= -\sigma_k^{-2}\left(\hat{\mathbf{H}}_k\check{\mathbf{B}}^{-1}\hat{\mathbf{H}}_k^H\check{\mathbf{G}}_k - \sum_{l=1}^{K}\left(\hat{\mathbf{H}}_k\check{\mathbf{B}}^{-1}\hat{\mathbf{H}}_l^H\check{\mathbf{G}}_l\check{\mathbf{G}}_l^H\hat{\mathbf{H}}_l\check{\mathbf{B}}^{-1}\hat{\mathbf{H}}_k^H\check{\mathbf{G}}_k\right)\right. \\ &\quad \left. - \sum_{l=1}^{K}\left(\text{tr}\left(\sigma_l^{-2}\check{\mathbf{G}}_l^H\hat{\mathbf{H}}_l\check{\mathbf{B}}^{-1}\left(\mathbf{R}_{\mathbf{H}_k}^{e,Tx}\right)^T\check{\mathbf{B}}^{-1}\hat{\mathbf{H}}_l^H\check{\mathbf{G}}_l\right)\right)\mathbf{R}_{\mathbf{H}_k}^{e,Rx}\check{\mathbf{G}}_k\right)\end{aligned}$$

and the projection operator is the simple solution by scaling all precoders with a factor $\sqrt{\tilde{P}/\sum_{k=1}^{K}\|\check{\mathbf{G}}_k\|_F^2}$. Finally, the uplink precoders $\bar{\mathbf{G}}_k, \forall\, k = 1,\ldots,K$ is obtained through the linear relation (5.77).

5.2.3.3 Outer Loop Optimization

Considering the outer-loop optimization (5.70), the problem can be similarly solved by a sub-gradient based method as (5.68). Herein, we also use the ellipsoid method to update μ_i, $\forall\ i = 0,\ldots,P$. Combined this with the inner loop optimization, the pseudo-code of the DB-DA is summarized in Algorithm 4.

5.2.4 Performance Comparison

First, the optimality and the complexity issues of the three proposed algorithms, the SOCP-based algorithm, the DL-DA, and the DB-DA, are briefly discussed as follows.

Optimality: As addressed before, the SOCP-based algorithm is guaranteed to converge to local optimum. Thus, we focus on the discussion of the latter two algorithms. First we consider the inner-loop optimization problems (5.52) and (5.69) in the DL-DA and the DB-DA, respectively. They are basically applying the constrained gradient projection method. Thus, the conditions under which the algorithm converges to

Algorithm 4 DB-DA robust design

Initialization: Set iteration number $l = 0$, maximal allowable number of iterations l_{max}, the desired accuracy ξ, the initial precoders $\check{\mathbf{G}}_k^{(0,0)}, \forall\, k = 1, \ldots, K$, and the ellipsoid center $\mu_i^{(0)}, \forall\, i = 0, \ldots, P$.
repeat
  $l \leftarrow l + 1$, for the fixed $\left\{\mu_i^{(l-1)}\right\}_{i=0}^{P}$.
  repeat
    $n \leftarrow n + 1$
    Update $\check{\mathbf{G}}_k^{(l,n)}$ via (5.80), $\forall\, k = 1, \ldots, K$.
  until MSE target converges.
  Using ellipsoid method to update $\mu_i^{(l)}, \forall\, p = 0, \ldots, P$.
until $\{\mu_i\}_{i=0}^{P}$ converges within the precision ξ.
Uplink to downlink conversion using (5.72).

the solution satisfying the first-order KKT optimality conditions has been addressed in [57, Theorem 1] and [15, Chapter 3]. If the optimization problem is convex, the solution is the global-optimum. Due to the favorable structure of the uplink problem, we find that the problem (5.69) is convex in the special case that all SUs have identical $\mathbf{R}_{\mathbf{H}_k}^{e,Tx}$ and $\mathbf{R}_{\mathbf{H}_k}^{e,Rx} = \mathbf{I}_{L_K}$, and they transmit as many data streams as the transmit antennas $N_k = L_k, \forall\, k = 1, \ldots, K$. Under such circumstances, the global optimum is achieved by the inner loop optimization in the DB-DA. Second, for the outer loop optimization, we use the diminishing step size rule for the sub-gradient-based method which guarantees the convergence to the optimal value [112].

Complexity: As observed in [57], the gradient projection algorithm exhibits a fast convergence rate when applied to the sum-MSE minimization problem, whereas the SOCP-based alternating optimization converges comparably slowly especially at high SNR values. Thus, we concentrate the discussion on the gradient projection method. The main computational complexity for each of them lies in the iterative update operations (5.62) and (5.80). After some simple analysis, both algorithms are shown to have the same computational order of the per-iteration operation. Therefore, the complexity depends on the number of iterations. In the numerical results, we compare the number of iterations of both algorithms for some given accuracy requirements. It is shown that the proposed approach requires less iterations, which validates the advantage of exploiting UL-DL duality.

Next, we evaluate the performance via simulation. Considering a cognitive network consisting of one BS and two SUs which share the same spectrum resource with two PUs, i.e., $K = 2$, $P = 2$, the simulation parameters are given as follows, except specified in each figure. The numbers of antennas at the BS, the SUs and the PUs are $N_t = 6$, $L_k = 2$ and $L_p = 1$, respectively. Two independent data streams are transmitted to a single SU, i.e., $N_k = 2$. Assuming the MMSE channel estimation method in [95] is performed, the covariance matrices of the channel error matrices are $\mathbf{R}_{\mathbf{H}_k}^{e,Rx} = \sigma_{es}^2 \mathbf{R}_{\mathbf{H}_k}^{Rx}$, $\mathbf{R}_{\mathbf{Y}_p}^{e,Rx} = \sigma_{ep}^2 \mathbf{R}_{\mathbf{Y}_p}^{Rx}$ and $\mathbf{R}_{\mathbf{H}_k}^{e,Tx} = \mathbf{R}_{\mathbf{Y}_p}^{e,Tx} = \mathbf{R}^{Tx}$, where σ_{es}^2 and σ_{ep}^2 are the

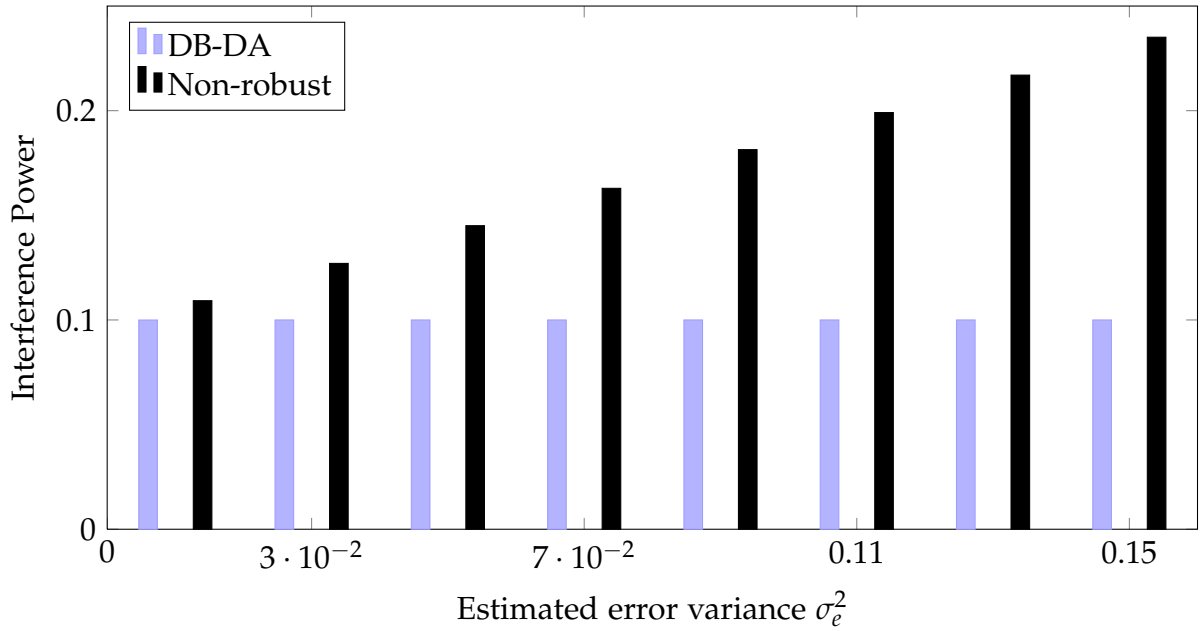

Figure 5.5: Average interference power vs. estimated error variance σ_e^2. For both SU links, SNR=10 dB.

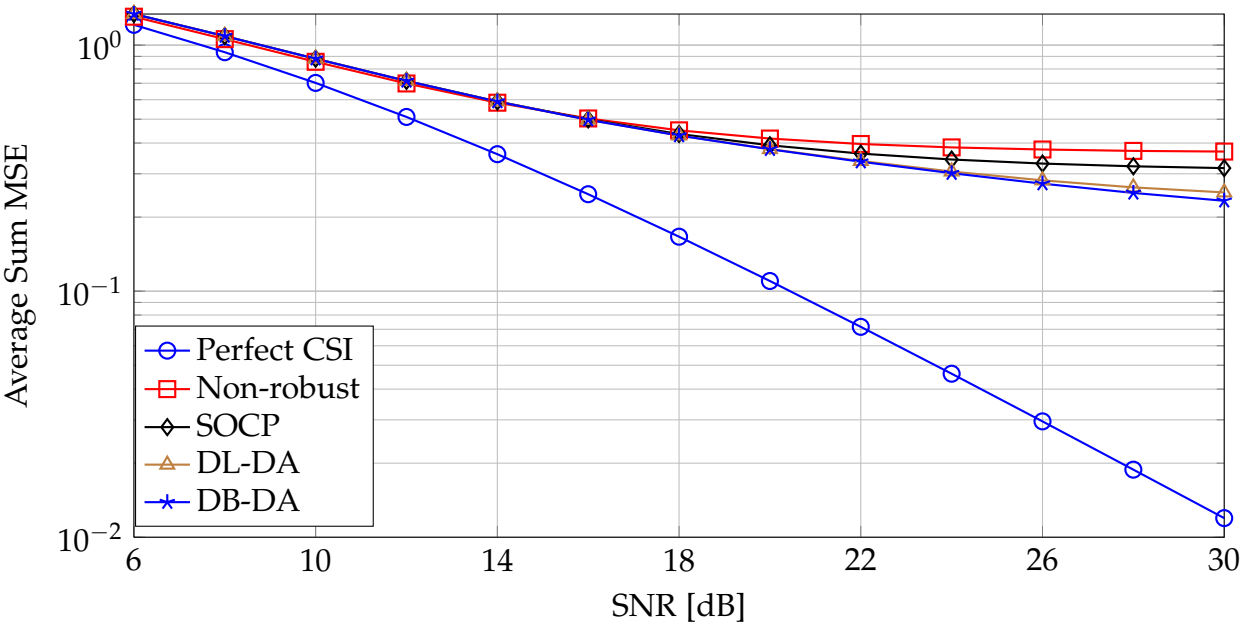

Figure 5.6: Average sum-MSE vs. SNR of SU links.

estimated error variances of the BS-SU links and the BS-PU links, respectively. The error variance is represented as σ_e^2 and $\sigma_e^2 = \sigma_{es}^2 = \sigma_{ep}^2 = 0.03$. The transmit covariance matrix $\mathbf{R}^{Tx}$ is modeled by a Toeplitz matrix with $\left[\mathbf{R}^{Tx}\right]_{i,j} = 0.1^{|i-j|}$. Similarly, the

Table 5.1: Average number of iterations vs. the desired accuracy ξ.

Accuracy requirement ξ	10^{-12}	10^{-10}	10^{-8}	10^{-6}	10^{-4}
DL-DA	250	223	172	81	27
DB-DA	24	18	12	8	4

entries $\left[\mathbf{R}_{\mathbf{H}_k}^{Rx}\right]_{i,j}$ and $\left[\mathbf{R}_{\mathbf{Y}_p}^{Rx}\right]_{i,j}$ are denoted by $0.2^{|i-j|}$. The accuracy of the algorithm is given by $\xi = 10^{-8}$. The transmit and interference power constraints are $P_0 = 0\,\text{dB}$ and $P_p = -10\,\text{dB}$, $\forall\, p = 1, \ldots, P$.

We compare the following approaches.

- *Perfect CSI*: the solution is based on the assumption that the available CSI at the secondary BS is accurate.
- *Non-robust*: the solution is based on the estimated CSI ignoring its imperfection.
- *SOCP*: proposed SOCP-based robust design in Algorithm 2.
- *DL-DA* : proposed downlink-based robust design in Algorithm 3.
- *DB-DA*: proposed duality-based robust design in Algorithm 4.

In the design of cognitive systems, the most important issue is to guarantee that performance degradation to the PUs is limited. Figure 5.5 plots the values of average interference power under the error variance σ_e^2 of channel estimation. The interference constraint is shown to be always violated by the non-robust design while strictly fulfilled by the robust solution, exemplified by the DB-DA. The violation gap increases with the larger variance due to the corresponding larger channel uncertainties.

Consider the MSE performance of the cognitive networks, Figure 5.6 demonstrates the average sum-MSE vs. SNR values. Both robust designs outperform the non-robust solution especially in the high SNR region. The performance gap seems small. The reason is that the a primary target of robust optimization is to guarantee the non-violation of the interference power constraint. Among the robust designs, the DB-DA achieves a marginal gain over the DL-DA, since the dual uplink problem has some favorable properties such as hidden convexity and less coupling of parameters, which yields the better convergence behavior. Both DB-DA and the DL-DA algorithms outperform the SOCP-based algorithm.

From the perspective of the convergence speed, we compare the average number of iterations of the DB-DA and the DL-DA method in Table 5.1. It is shown that the DB-DA requires less number of iterations for given desired accuracies, i.e., has faster convergence speed. As analyzed before, the complexity per-iteration of both methods are of the same computational order. Therefore, the DB-DA has lower complexity.

5.3 Summary

Exploiting multiple antennas for the linear transceiver optimization is considered as an effective way to mitigate the interference in the underlay paradigm of a cognitive system. However, the performance can be deteriorated by imperfect CSI since the non-robust design neglecting the channel uncertainties leads to the violation of the interference constraint and thus limits the application of cognitive systems. Two kinds of CSI error models are therefore commonly assumed: the bounded and the stochastic model. In this chapter, we have studied the robust transceiver optimization considering both models. Specifically, on the one hand, assuming the CSI error is bounded within an ellipsoidal region, we employed the worst-case principle to minimize the maximum per-user MSE for any channel realization within the CSI uncertainty region. An alternating algorithm was used to optimize the precoding and equalizer filters iteratively until convergence to a local optimum is achieved. Each subproblem was reformulated into an equivalent convex SDP form. On the other hand, if the CSI error follows the stochastic model, the optimization aims at minimizing the sum-MSE of the secondary network averaged over all channel realizations within the uncertainty region. Three robust algorithms were proposed: The first one applies the alternating principle to optimize the precoding and equalizer filters iteratively. Each subproblem was converted into the convex SOCP form. The second one, the DL-DA algorithm, utilized the gradient projection method [15] in the downlink transmission. The third one, the DB-DA algorithm, performed the optimization in the dual uplink based on the established MSE duality incorporating multiple power constraints and CSI imperfection. We also addressed the complexity and optimality issues of the proposed methods. The effectiveness and convergence of the algorithms were validated and compared by numerical results. The advantage of exploiting the UL–DL duality for the downlink optimization is verified, in terms of both complexity and optimality.

Chapter 6

Sensing-Based Power Allocation

Up to now, we have investigated the optimization of spectrum sensing and transceiver strategies in the interweave and the underlay paradigm, respectively. Recent works have revealed that a significant improvement in the spectrum efficiency is obtained by joint optimization of sensing, access, and transmission strategies [8, 21, 22, 70, 91, 97, 100–102, 107, 117, 119]. This can be considered as a hybrid paradigm, i.e., a combination of the above-mentioned two paradigms.[1] In this chapter, we apply this idea to the system considered in Chapter 4 by additionally allowing the SU to consider the reliability of the sensing outcome and adjust its transmission strategy accordingly. Both soft-decision and hard-decision sensing results are considered. The purpose is to characterize the gain achieved by the SUs through the effective utilization of the sensing information.

In practice, the SUs only have partial CSI related to the primary link. Since such CSI may highly depend on the location information [61], the effect of location uncertainty [125] is also considered in modeling the interference caused by the SUs to the PUs. This idea shares some similarity with [130] where the primary exclusive region is determined based on the interference generated by the SUs considering the randomness of the users' deployment. However, we are not aiming at characterizing such spatial opportunities for the SUs, instead we want to assess the interference level resulting from the secondary transmission.

In what follows, we design power allocation strategies in the hybrid paradigm. After introducing the system model in Section 6.1, the cases that the sensing metric is in a general form or a special form of a binary hard decision are investigated in Section 6.2 and Section 6.3, respectively. Specific attention is paid to the hard-decision sensing based scheme because it can be directly compared with the interweave and underlay paradigms. In addition, we consider the location uncertainty of the primary network and provide a model of interference caused by the secondary transmission in Section 6.4. The result is then incorporated into the optimization problem to design the transmission strategy.

The results presented in this chapter are addressed in part by the author in [51] for possible future publication[2].

[1] In some works, e.g., [152], the term *overlay* corresponds to what we term *interweave* in this work.

[2] In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of RWTH Aachen University's products or services. Internal or personal use of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution, please go to `http://www.ieee.org/publications_standards/publications/rights/rights_link.html` to learn how to obtain a License from RightsLink.

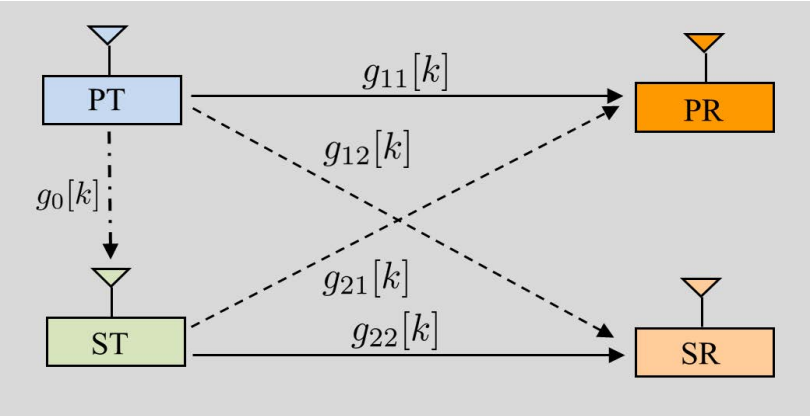

Figure 6.1: System model

6.1 System Model

The considered system is depicted in Figure 6.1. Compared to Figure 4.1, we additionally introduce the sensing link from the PT to the ST with channel power gain denoted by $g_0[k]$ at the kth time instant. The remaining notations and the input-output relation are described in Section 4.1.

Each secondary transmission process is divided into two phases: the spectrum sensing phase and data transmission phase with the duration given by T_{se} and T_{tran}, respectively. The total duration follows as

$$T_{\text{all}} = T_{\text{se}} + T_{\text{tran}}.$$

In the first phase, the ST calculates the sensing metric γ based on the observations received from the PT through the sensing link. In the second phase, the secondary transmit power is adapted according to γ and the available CSI. Herein, we do not restrict the discussion to any specific sensing algorithm. Therefore, γ stands for any sensing metric such as the signal energy, sampled correlation matrix of the received signals, or a binary hard decision.

Due to limited cooperation between the PUs and the SUs, the instantaneous CSI related to the primary links is difficult to obtain at the ST. Hence, we make the same CSI assumptions as in Section 4.2.2. For the remainder of the chapter, we omit the time argument for simplicity. Specifically, statistical channel parameters l_{12} and l_{21} of the PT-SR and the ST-PR link are available at the ST. Additionally, the ST has the knowledge of instantaneous CSI g_{22}. The SR is assumed to know instantaneous CSI g_{12} and g_{22}.

6.2 Soft-Decision Sensing

During the sensing phase, the ST gathers the observation received from the PT and calculates the sensing measures γ. Moreover, the ST knows the probability of the

presence and absence of the primary transmission, named as "ON" and "OFF" statuses, i.e., $p(\mathcal{H}_1)$ and $p(\mathcal{H}_0)$, respectively. If γ is a continuous variable, the PDF of γ conditioned on the status "ON" and "OFF" is given as $f(\gamma|\mathcal{H}_1)$ and $f(\gamma|\mathcal{H}_0)$, respectively; If γ is a discrete variable, we use the same notation $f(\gamma|\mathcal{H}_i)$ to denote the probability mass function (PMF) of γ conditioned on $\mathcal{H}_i$, $\forall\, i = 1, 2$. These distribution parameters $f(\gamma|\mathcal{H}_i)$, $\forall\, i = 1, 2$, are also known at the ST.

Consequently, given a certain γ, the probability that the primary transmission is present and absent can be calculated using the Bayes' theorem as

$$p(\mathcal{H}_0|\gamma) = \frac{p(\mathcal{H}_0) f(\gamma|\mathcal{H}_0)}{f(\gamma)} \tag{6.1}$$

$$p(\mathcal{H}_1|\gamma) = 1 - p(\mathcal{H}_0|\gamma) \tag{6.2}$$

respectively, where $f(\gamma)$ is the PDF of γ with

$$f(\gamma) = p(\mathcal{H}_0) f(\gamma|\mathcal{H}_0) + p(\mathcal{H}_1) f(\gamma|\mathcal{H}_1).$$

Here, we only consider the scenario in which $f(\gamma)$ is non-zero, i.e., $f(\gamma) > 0$.

Two power constraints are imposed. First, we limit the average interference power received at the PR by P_I if the primary transmission is active. Second, the transmit power of the ST is restricted to P_S. Both P_I and P_S are positive and finite. We aim at optimizing the achievable rate of the secondary link based on the sensing metric γ and the available CSI. Thus, introducing the optimization variable $P_2(\gamma, g_{22})$ as the power level adapted to the sensing measure γ and instantaneous CSI g_{22}, the optimization problem is formulated as

$$\begin{aligned}
\max_{P_2(\gamma,g_{22})} \quad & \mathbb{E}_{\gamma,g_{22}}\left\{ p(\mathcal{H}_0|\gamma) \ln\left(1 + \frac{P_2(\gamma,g_{22}) g_{22}}{\sigma_s^2}\right)\right\} \\
& + \mathbb{E}_{\gamma,g_{12},g_{22}}\left\{ p(\mathcal{H}_1|\gamma) \ln\left(1 + \frac{P_2(\gamma,g_{22}) g_{22}}{P_1 g_{12} + \sigma_s^2}\right)\right\} \\
\text{s.t.} \quad & P_2(\gamma,g_{22}) \le P_S, \qquad \forall g_{22} \ge 0 \\
& P_2(\gamma,g_{22}) \ge 0, \qquad \forall g_{22} \ge 0 \\
& \mathbb{E}_{\gamma,g_{22},g_{21}}\{P_2(\gamma,g_{22}) g_{21}|\mathcal{H}_1\} \le P_I.
\end{aligned} \tag{6.3}$$

Using the mutual independence among the CSI of different links, e.g., g_{22} and g_{21}, the last constraint is converted into

$$\mathbb{E}_{\gamma,g_{22}}\{P_2(\gamma,g_{22})|\mathcal{H}_1\} \le P_I l_{21} \tag{6.4}$$

which has the similar form as an average transmit power constraint.

6.2.1 Optimal Power Allocation

Because the expectation and summation operations preserve convexity [115], the objective function in optimization problem (6.3) is a continuously differentiable concave

function in nonnegative variable $P_2(\gamma, g_{22})$. Moreover, all the constraints are continuously differentiable and affine functions of $P_2(\gamma, g_{22})$. Hence, the optimization problem (6.3) is a convex optimization problem. A regularity condition, e.g., the Slater condition, is also satisfied. Consequently, the KKT conditions are the necessary and sufficient optimality conditions. In the following, we apply the KKT conditions in function space to this optimization problem and derive the optimal solution.

We first express the objective function in (6.3) as an explicit function w.r.t. the available CSI. Similar to (4.32), the objective function is averaged over different channel realizations of g_{12}, it yield

$$\begin{aligned}&\mathbb{E}_{g_{12},g_{22}}\left\{\ln\left(1+\frac{P_2(\gamma,g_{22})g_{22}}{\sigma_s^2+P_1g_{12}}\right)\right\}\\&=\mathbb{E}_{g_{22}}\left\{\ln\left(1+\frac{P_2(\gamma,g_{22})g_{22}}{\sigma_s^2}\right)+e^{s_1}E_1(s_1)\right\}-e^{s_0}E_1(s_0)\end{aligned}\tag{6.5}$$

where the terms s_0 and s_1 are introduced for the sake of brevity

$$\begin{aligned}s_0&=\frac{l_{12}\sigma_s^2}{P_1}\\s_1&=\frac{l_{12}\left(\sigma_s^2+P_2(\gamma,g_{22})g_{22}\right)}{P_1}.\end{aligned}$$

Then, the Lagrangian of (6.3) is

$$\begin{aligned}L=&\mathbb{E}_{\gamma,g_{22}}\left\{p(\mathcal{H}_0|\gamma)\ln\left(1+\frac{P_2(\gamma,g_{22})g_{22}}{\sigma_s^2}\right)\right\}\\&+\mathbb{E}_{\gamma,g_{22}}\left\{p(\mathcal{H}_1|\gamma)\left(\ln\left(1+\frac{P_2(\gamma,g_{22})g_{22}}{\sigma_s^2}\right)+e^{s_1}E_1(s_1)-e^{s_0}E_1(s_0)\right)\right\}\\&-\lambda_2\left(\mathbb{E}_{\gamma,g_{22}}\left\{P_2\left(\gamma,g_{22}\right)|\mathcal{H}_1\right\}-P_Il_{21}\right)\end{aligned}\tag{6.6}$$

where λ_2 is the non-negative Lagrangian multiplier corresponding to the constraint (6.4). Applying the KKT conditions for functional optimization [87] to (6.3), the optimal solution $P_2^\star(\gamma, g_{22})$ satisfies

$$\left.\frac{dl}{dP_2(\gamma,g_{22})}\right|_{P_2(\gamma,g_{22})=P_2^\star(\gamma,g_{22})}\begin{cases}\geq 0, & P_2^\star(\gamma,g_{22})=P_S\\ \leq 0, & P_2^\star(\gamma,g_{22})=0\\ =0, & 0<P_2^\star(\gamma,g_{22})<P_S\end{cases}\tag{6.7}$$

$$\lambda_2^\star\geq 0\tag{6.8}$$

$$\lambda_2^\star\left(\mathbb{E}_{\gamma,g_{22}}\left\{P_2(\gamma,g_{22})|\mathcal{H}_1\right\}-P_Il_{21}\right)=0\tag{6.9}$$

$$\mathbb{E}_{\gamma,g_{22}}\left\{P_2(\gamma,g_{22})|\mathcal{H}_1\right\}\leq P_Il_{21}\tag{6.10}$$

where the function l is defined as

$$\begin{aligned} l = & p(\mathcal{H}_0|\gamma) \ln\left(1+\frac{P_2(\gamma,g_{22})g_{22}}{\sigma_s^2}\right) \\ & + p(\mathcal{H}_1|\gamma)\left(\ln\left(1+\frac{P_2(\gamma,g_{22})g_{22}}{\sigma_s^2}\right)+e^{s_1}E_1(s_1)-e^{s_0}E_1(s_0)\right) \\ & - \lambda_2\left(\frac{p\left(\mathcal{H}_1|\gamma\right)}{p(\mathcal{H}_1)}P_2(\gamma,g_{22})-P_l l_{21}\right). \end{aligned} \tag{6.11}$$

and the derivative on the left hand side of (6.7) is

$$\begin{aligned} \frac{dl}{dP_2(\gamma,g_{22})} = & \left(p(\mathcal{H}_0|\gamma)+p(\mathcal{H}_1|\gamma)\right)\frac{1}{\sigma_s^2/g_{22}+P_2(\gamma,g_{22})} \\ & + p(\mathcal{H}_1|\gamma)\left(\frac{l_{12}g_{22}}{P_1}e^{s_1}E_1(s_1)-\frac{1}{\sigma_s^2/g_{22}+P_2(\gamma,g_{22})}\right) \\ & - \lambda_2^\star\frac{p\left(\mathcal{H}_1|\gamma\right)}{p(\mathcal{H}_1)}. \end{aligned} \tag{6.12}$$

We multiply both sides of (6.12) with positive $f(\gamma)$ and insert (6.1) and (6.2) into it. After some mathematical manipulations, we obtain $P_2^\star(\gamma,g_{22})$ from solving KKT conditions

$$P_2^\star(\gamma,g_{22}) = \begin{cases} 0, & H_2(\alpha_{0,\mathrm{h}},\alpha_{1,\mathrm{h}},0) \le \lambda_2^\star f(\gamma|\mathcal{H}_1) \\ p^\star(\gamma,g_{22}), & H_2(\alpha_{0,\mathrm{h}},\alpha_{1,\mathrm{h}},P_S) < \lambda_2^\star f(\gamma|\mathcal{H}_1) < H_2(\alpha_{0,\mathrm{h}},\alpha_{1,\mathrm{h}},0) \\ P_S, & H_2(\alpha_{0,\mathrm{h}},\alpha_{1,\mathrm{h}},P_S) \ge \lambda_2^\star f(\gamma|\mathcal{H}_1). \end{cases} \tag{6.13}$$

with $H_2(\alpha_0,\alpha_1,x)$ given as

$$H_2(\alpha_0,\alpha_1,x) = \frac{\alpha_0 P_1}{l_{12}\left(\sigma_s^2+xg_{22}\right)} + \alpha_1 e^{\frac{l_{12}\left(\sigma_s^2+xg_{22}\right)}{P_1}} E_1\left(\frac{l_{12}\left(\sigma_s^2+xg_{22}\right)}{P_1}\right) \tag{6.14}$$

and

$$\begin{aligned} \alpha_{0,\mathrm{h}} &= p(\mathcal{H}_0)f(\gamma|\mathcal{H}_0)\frac{l_{12}g_{22}}{P_1} \\ \alpha_{1,\mathrm{h}} &= p(\mathcal{H}_1)f(\gamma|\mathcal{H}_1)\frac{l_{12}g_{22}}{P_1}. \end{aligned}$$

In (6.13), the value $p^\star(\gamma,g_{22})$ is equal to the solution x of the following equation

$$H_2(\alpha_{0,\mathrm{h}},\alpha_{1,\mathrm{h}},x) = \lambda_2^\star f(\gamma|\mathcal{H}_1). \tag{6.15}$$

Using the result in Appendix A.4, it is easy to verify that $H_2(\alpha_{0,\mathrm{h}},\alpha_{1,\mathrm{h}},x)$ is strictly monotonically decreasing on x, $x \in (0,P_S)$, for finite and positive g_{22}. Therefore, the bisection method can be applied to search for the root of (6.15).

The variable $\lambda_2^\star$ is determined to satisfy the constraint (6.4), similar in the supervised thesis [104]. More particularly, If $P_S \leq P_I l_{21}$, the ST can always transmit with power P_S while the average interference power constraint is strictly satisfied. Consequently, the computation of $\lambda_2^\star$ is not required. If $P_S > P_I l_{21}$, the optimal $\lambda_2^\star$ should be chosen to let the inequality (6.4) be satisfied with equality. We remark that λ_2^* is finite and positive in this case. Specifically, first, if $\lambda_2^* = 0$ holds, then the optimal solution (6.13) results $P_2^*(\gamma, g_{22}) = P_S$ for any positive g_{22} by noting that $H_2(\alpha_{0,\mathrm{h}}, \alpha_{1,\mathrm{h}}, x)$ in (6.14) has positive value for non-negative x and finite g_{22}. This violates the average power contraint since the average power under such circumstance is equal to P_S and larger than $P_I l_{21}$. Second, if λ_2^* is infinitely large, this yields $P_2^*(\gamma, g_{22}) = 0$ which is obviously not optimal.

Based on the above discussion, the optimal solution for the case $P_S \leq P_I l_{21}$ is $P_2^*(\gamma, g_{22}) = P_S$, which can be straightforwardly obtained. Therefore, we focus on the solution for the case $P_S > P_I l_{21}$ in which λ_2^* is finite and positive.

6.2.2 Suboptimal Power Allocation

Revising the optimal strategy, the relationship between the solution (6.13) and the effect of the CSI is not explicitly shown. Moreover, a real-time numerical calculation of $p^\star(\gamma, g_{22})$ is required for each g_{22} within its feasible region. Thus, the optimal strategy is time-consuming. In this subsection, we propose a low complexity suboptimal power allocation. The relation between the power level and the CSI is explicitly described.

Motivated by the effectiveness of the suboptimal strategy in Section 4.2.2.2, we use the following approximation to simplify the optimal solution

$$e^x E_1(x) \approx \frac{1}{x+1}. \tag{6.16}$$

Specifically, the first condition in (6.13), $H_2(\alpha_{0,\mathrm{h}}, \alpha_{1,\mathrm{h}}, 0) \leq \lambda_2^\star p(\gamma|\mathcal{H}_1)$, is approximated by

$$g_{22} \leq \frac{\lambda_2^\star f(\gamma|\mathcal{H}_1)}{p(\mathcal{H}_0) f(\gamma|\mathcal{H}_0)/\sigma_s^2 + p(\mathcal{H}_1) f(\gamma|\mathcal{H}_1)/(\sigma_s^2 + P_1/l_{12})} \tag{6.17}$$

which shows that the optimal strategy is to switch off the transmission when g_{22} is smaller than a threshold g_{low} with

$$g_{\mathrm{low}} = \frac{\lambda_2^\star f(\gamma|\mathcal{H}_1)}{p(\mathcal{H}_0) f(\gamma|\mathcal{H}_0)/\sigma_s^2 + p(\mathcal{H}_1) f(\gamma|\mathcal{H}_1)/(\sigma_s^2 + P_1/l_{12})}. \tag{6.18}$$

Similarly, the condition $H_2(\alpha_{0,\mathrm{h}}, \alpha_{1,\mathrm{h}}, P_S) \geq \lambda_2^\star f(\gamma|\mathcal{H}_1)$ is approximated by

$$\frac{p(\mathcal{H}_1) f(\gamma|\mathcal{H}_1)}{(\sigma_s^2 + P_1/l_{12})/g_{22} + P_S} + \frac{p(\mathcal{H}_0) f(\gamma|\mathcal{H}_0)}{\sigma_s^2/g_{22} + P_S} \geq \lambda_2^\star f(\gamma|\mathcal{H}_1) \tag{6.19}$$

in which the left hand side is a monotonically increasing function of g_{22}. Hence, there exists a g_{upp} such that $g_{22} \geq g_{\mathrm{upp}}$ indicates the satisfaction of the condition (6.19). The

solution of g_{upp} is calculated in Appendix A.17. Applying the approximation (6.16) to (6.15), we obtain an approximation of $p^\star(\gamma, g_{22})$ as

$$\tilde{p}^\star(\gamma, g_{22}) = \frac{\tilde{x}P_1/l_{12} - \sigma_s^2}{g_{22}} \tag{6.20}$$

where $\tilde{x}$ is calculated by a similar method as discussed in Appendix A.17:

$$\tilde{x} = \frac{\alpha_{0,\text{h}} + \alpha_{1,\text{h}} - \lambda_2^\star f(\gamma|\mathcal{H}_1) + \sqrt{\left(\alpha_{0,\text{h}} + \alpha_{1,\text{h}} - \lambda_2^\star f(\gamma|\mathcal{H}_1)\right)^2 + 4\alpha_{0,\text{h}}\lambda_2^\star f(\gamma|\mathcal{H}_1)}}{2\lambda_2^\star f(\gamma|\mathcal{H}_1)}. \tag{6.21}$$

In summary, the suboptimal power control strategy is

$$\tilde{P}_2^\star(\gamma, g_{22}) = \begin{cases} 0, & g_{22} \leq g_{\text{low}} \\ \tilde{p}^\star(\gamma, g_{22}), & g_{\text{low}} < g_{22} < g_{\text{upp}} \\ P_S, & \text{otherwise.} \end{cases} \tag{6.22}$$

6.3 Hard-Decision Sensing

Hard-decision sensing-based power control is encompassed as a special case of soft-decision sensing-based scheme considered in Section 6.2. Specifically, the sensing measure γ is reduced to binary statuses $\hat{\mathcal{H}}_0$ and $\hat{\mathcal{H}}_1$ representing the declaration of the primary transmission as present or absent, respectively. However, a specific investigation on this problem is still informative since it can be directly compared to the standard underlay and interweave paradigms. Thus, the result of this problem serves the purpose to exemplify the performance gain over the reference paradigms. In this section, we address the optimal and suboptimal power allocation solutions based on hard-decision sensing. Note that the notation $f(\cdot)$ indicates a PMF. Due to the close relation to Section 6.2, we omit the detailed derivations and present the final results.

6.3.1 Optimal and Suboptimal Power Allocation

Simplifying the sensing metric γ into two statuses $\hat{\mathcal{H}}_0$ and $\hat{\mathcal{H}}_1$, we have

$$\gamma = \hat{\mathcal{H}}_0 \;\Rightarrow\; \begin{cases} f(\gamma|\mathcal{H}_0) = 1 - P_{FA} \\ f(\gamma|\mathcal{H}_1) = 1 - P_D \end{cases} \tag{6.23}$$

$$\gamma = \hat{\mathcal{H}}_1 \;\Rightarrow\; \begin{cases} f(\gamma|\mathcal{H}_0) = P_{FA} \\ f(\gamma|\mathcal{H}_1) = P_D \end{cases} \tag{6.24}$$

Inserting them into the optimization problem (6.3), it is simplified to

$$\max_{P_{2,0}(g_{22}),\, P_{2,1}(g_{22})} \quad R_{\text{hd}} \tag{6.25}$$

$$\begin{aligned}\text{s.t. } & P_{2,0}(g_{22}) \le P_S,\ P_{2,1}(g_{22}) \le P_S \qquad \forall g_{22} \ge 0 \\ & P_{2,0}(g_{22}) \ge 0,\quad P_{2,1}(g_{22}) \ge 0, \qquad \forall g_{22} \ge 0 \\ & \mathbb{E}_{g_{22}}\left\{(1-P_D)P_{2,0}(g_{22}) + P_D P_{2,1}(g_{22})\right\} \le P_I l_{21}\end{aligned}$$

where the objective function is

$$\begin{aligned}R_{\text{hd}} = \ & \mathbb{E}_{g_{22}}\left\{p(\mathcal{H}_0)\left((1-P_{FA})\ln\left(1+\frac{P_{2,0}(g_{22})g_{22}}{\sigma_s^2}\right)+P_{FA}\ln\left(1+\frac{P_{2,1}(g_{22})g_{22}}{\sigma_s^2}\right)\right)\right\} \\ & + \mathbb{E}_{g_{12},g_{22}}\left\{p(\mathcal{H}_1)\left((1-P_D)\ln\left(1+\frac{P_{2,0}(g_{22})g_{22}}{P_1 g_{12}+\sigma_s^2}\right)\right.\right. \\ & \left.\left.+P_D\ln\left(1+\frac{P_{2,1}(g_{22})g_{22}}{P_1 g_{12}+\sigma_s^2}\right)\right)\right\}. \qquad (6.26)\end{aligned}$$

Since the problem (6.25) is the special case of the problem (6.3) based on soft decisions, the optimal solutions are derived based on (6.13) by incorporating the PMFs (6.23) and (6.24). We directly present the results as follows:

$$P_{2,0}^{\star}(g_{22}) = \begin{cases}0, & H_2(\alpha_{00,\text{hd}},\alpha_{10,\text{hd}},0) \le \lambda_2^{\star}(1-P_D) \\ p_0^{\star}(g_{22}), & H_2(\alpha_{00,\text{hd}},\alpha_{10,\text{hd}},P_S) < \lambda_2^{\star}(1-P_D) < H_2(\alpha_{00,\text{hd}},\alpha_{10,\text{hd}},0) \\ P_S, & \text{otherwise}\end{cases} \qquad (6.27)$$

$$P_{2,1}^{\star}(g_{22}) = \begin{cases}0, & H_2(\alpha_{01,\text{hd}},\alpha_{11,\text{hd}},0) \le \lambda_2^{\star}P_D \\ p_1^{\star}(g_{22}), & H_2(\alpha_{01,\text{hd}},\alpha_{11,\text{hd}},P_S) < \lambda_2^{\star}P_D < H_2(\alpha_{01,\text{hd}},\alpha_{11,\text{hd}},0) \\ P_S, & \text{otherwise.}\end{cases} \qquad (6.28)$$

with

$$\begin{aligned}\alpha_{00,\text{hd}} &= p(\mathcal{H}_0)(1-P_{FA})\frac{l_{12}g_{22}}{P_1} \\ \alpha_{10,\text{hd}} &= p(\mathcal{H}_1)(1-P_D)\frac{l_{12}g_{22}}{P_1}. \\ \alpha_{01,\text{hd}} &= p(\mathcal{H}_0)P_{FA}\frac{l_{12}g_{22}}{P_1} \\ \alpha_{11,\text{hd}} &= p(\mathcal{H}_1)P_D\frac{l_{12}g_{22}}{P_1}.\end{aligned}$$

The notations $P_{2,0}^{\star}(g_{22})$ and $P_{2,1}^{\star}(g_{22})$ represent the optimal power control under the sensing decision $\hat{\mathcal{H}}_0$ and $\hat{\mathcal{H}}_1$, respectively. The intermediate power values $p_0^{\star}(g_{22})$ and $p_1^{\star}(g_{22})$ can obtained as the roots of the following equations

$$H_2\left(\alpha_{00,\text{hd}},\alpha_{10,\text{hd}},p_0^{\star}(g_{22})\right) = \lambda_2^{\star}(1-P_D) \qquad (6.29)$$

$$H_2\left(\alpha_{01,\text{hd}},\alpha_{11,\text{hd}},p_1^{\star}(g_{22})\right) = \lambda_2^{\star}P_D \qquad (6.30)$$

respectively, through some numerical method such as the bisection method.

The efficient suboptimal strategy can be derived analogously to (6.22) by incorporating the PMFs (6.23) and (6.24).

6.3.2 Referenced Strategies

In order to provide a thorough comparison between the different standard cognitive radio paradigms, the development of power allocation strategies is required for the standard paradigms. In this subsection, we show that the power optimization problem in both interweave and underlay paradigm can be formulated as a special case of the sensing-based power control in Section 6.3.1. Hence, the methodology for the derivation of the optimal and suboptimal power allocation strategies in Section 6.3.1 can be straightforwardly applied.

6.3.2.1 Interweave Paradigm

In the interweave paradigm the SUs opportunistically access the spectrum only if they declare the inactivity of the primary transmission. Thus, the optimization problem is formulated as

$$\begin{aligned} \max_{P_{\text{in}}(g_{22})} \quad & R_{\text{in}} \\ \text{s.t.} \quad & P_{\text{in}}(g_{22}) \leq P_S, \qquad \forall g_{22} \geq 0 \\ & P_{\text{in}}(g_{22}) \geq 0, \qquad \forall g_{22} \geq 0 \\ & (1-P_D)\mathbb{E}_{g_{22}}\left\{P_{\text{in}}(g_{22})\right\} \leq P_I l_{21} \end{aligned} \tag{6.31}$$

with

$$\begin{aligned} R_{\text{in}} = \mathbb{E}_{g_{12},g_{22}}\Bigg\{ & p(\mathcal{H}_0)(1-P_{FA})\ln\left(1+\frac{P_{\text{in}}(g_{22})g_{22}}{\sigma_s^2}\right) \\ & + p(\mathcal{H}_1)(1-P_D)\ln\left(1+\frac{P_{\text{in}}(g_{22})g_{22}}{P_1 g_{12}+\sigma_s^2}\right)\Bigg\} \end{aligned} \tag{6.32}$$

which is a special case of (6.26) by setting $P_{2,0}(g_{22}) = P_{\text{in}}(g_{22})$ and $P_{2,1}(g_{22}) = 0$, $\forall\, g_{22} \geq 0$, i.e., the ST only transmits under the decision $\mathcal{H}_0$. The power allocation strategies to solve (6.33) can be straightforwardly derived from Section 6.3.1.

6.3.2.2 Underlay Paradigm

In the underlay paradigm, the power allocation strategy is designed assuming the primary transmission is always active. Consequently, the optimization problem is

$$\begin{aligned} \max_{P_{\text{un}}(g_{22})} \quad & R_{\text{un}} \\ \text{s.t.} \quad & P_{\text{un}}(g_{22}) \leq P_S, \qquad \forall g_{22} \geq 0 \end{aligned} \tag{6.33}$$

$$P_{\text{un}}(g_{22}) \geq 0, \qquad \forall g_{22} \geq 0$$
$$\mathbb{E}_{g_{22}}\{P_{\text{un}}(g_{22})\} \leq P_I l_{21}$$

with

$$R_{\text{un}} = \mathbb{E}_{g_{12},g_{22}}\left\{p(\mathcal{H}_0)\ln\left(1+\frac{P_{\text{un}}(g_{22})g_{22}}{\sigma_s^2}\right)+p(\mathcal{H}_1)\ln\left(1+\frac{P_{\text{un}}(g_{22})g_{22}}{P_1 g_{12}+\sigma_s^2}\right)\right\}. \quad (6.34)$$

We note that (6.34) is a special case of (6.26) by setting $P_{2,0}(g_{22}) = P_{2,1}(g_{22}), \forall\, g_{22} \geq 0$, i.e., the power level is not adapted to the sensing observations. The power allocation strategies to solve (6.33) can be directly obtained from the results in Section 6.3.1.

For the evaluation of the proposed strategies, we set the simulation parameters as follows. The statistical CSI parameter is $l_{22} = 1$, $l_{21} = 1$, and $l_{12} = 2$. The noise variance of the secondary link is $\sigma_s^2 = 1$. All power values in dB are given relative to the reference level of value 1. The transmit power of the PT is 10 dB. The interference power limit P_I is chosen between three values -5, 0, and 5 dB. The probabilities that the primary transmission is active and inactive are equal, i.e., $p(\mathcal{H}_0) = p(\mathcal{H}_1) = 0.5$. According to the requirements of IEEE 802.22 WRAN, the performance of spectrum sensing should guarantee the false alarm rate below 10% and a probability of detection above 90%. Herein, the ST uses energy detection as the sensing algorithm with the observation length equal to 30. We choose the threshold on the ROC curve of energy detection that it yields $P_{\text{FA}} = 9.84\%$ and $P_{\text{D}} = 92.32\%$.

Power allocation strategies for different paradigms are compared. They are listed as follows.

- Optimal algorithm for hybrid paradigm based on hard-sensing decisions: the solution is given in (6.27) and (6.28).
- Suboptimal algorithm for hybrid paradigm based on hard-sensing decisions: the power value is in the form of (6.22) by incorporating (6.23) and (6.24).
- Suboptimal algorithm for the interweave paradigm: the solution is similar to that of the hybrid paradigm as discussed in Section 6.3.2.1.
- Suboptimal algorithm for underlay paradigm: the solution is similar to that of the hybrid paradigm as elaborated in Section 6.3.2.2.

Note that in order to reduce the computational complexity, only the suboptimal strategies in the reference paradigms are considered. This idea is based on the observation that the near-optimality of the suboptimal power allocation strategies is achieved for the hybrid paradigm, as evidenced by the following numerical results.

Figure 6.2 shows the achievable rate of the secondary link vs. the ratio $\rho = P_S/P_I$. We only consider the case $\rho > 1$ since the interference power constraint is usually stricter than the transmit power constraint. The detailed explanations of the plots are given below the figure. The near-optimality of the suboptimal algorithm is validated for the hybrid paradigm. Moreover, the power adaptation in the underlay paradigm suffers from a great performance loss compared to the hybrid one due to the lack of

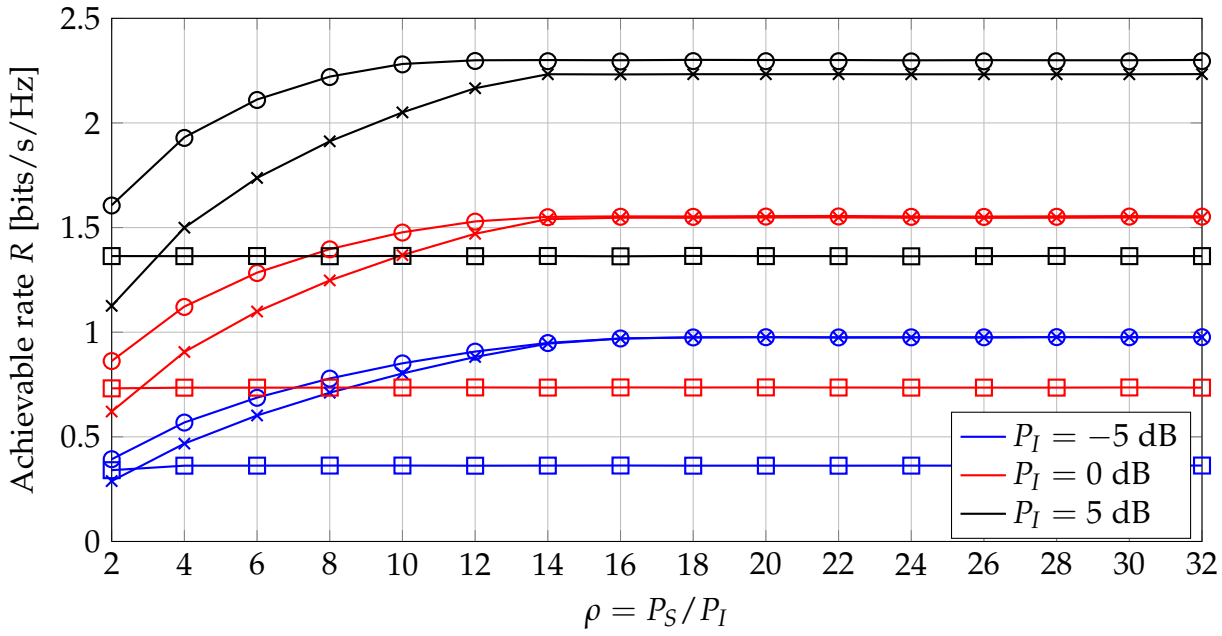

Figure 6.2: Achievable rate of the secondary link vs. the ratio $\rho = P_S/P_I$. Solid lines: optimal algorithm for the hybrid paradigm based on hard-sensing decisions; circles: suboptimal algorithm for the hybrid paradigm based on hard-sensing decisions; solid lines marked with crossings: suboptimal algorithm for the interweave paradigm; solid lines marked with squares: suboptimal algorithm for the underlay paradigm.

sensing observations. Finally, the hybrid paradigm also outperforms the interweave one due to the additional transmission during the cases even the presence of the primary transmission is detected.

Remarks: Strictly speaking, the algorithm for the underlay paradigm is not evaluated in a "fair" manner compared with other paradigms. Specifically, the actual transmission process of the underlay paradigm only requires the data transmission phase, which is in contrary with the hybrid and interweave ones which require the two phases protocol, i.e., spectrum sensing and data transmission phases during each process. A fairer comparison is to scale its achievable performance of the underlay paradigm with the factor $T_{\text{all}}/T_{\text{tran}}$, where T_{all} and T_{tran} are the duration of two phases and data transmission phase, respectively. Due to the fact that the duration of two phases is larger than a sole phase, i.e., $T_{\text{all}}/T_{\text{tran}} > 1$, the performance of the underlay paradigm shown here is the lower-bound of the actual performance. However, such duration lengths are not clarified in the current standard and the study on this issue is still ongoing. Hence, in order to keep consistent with the other paradigms, we assume here the underlay paradigm also employs the two phases process but with the sensing phase wasted for usage. If the sensing phase is short enough compared with the data transmission phase, i.e., $T_{\text{all}}/T_{\text{tran}} \approx 1$, herein the performance of the underlay paradigm asymptotically represent the actual performance.

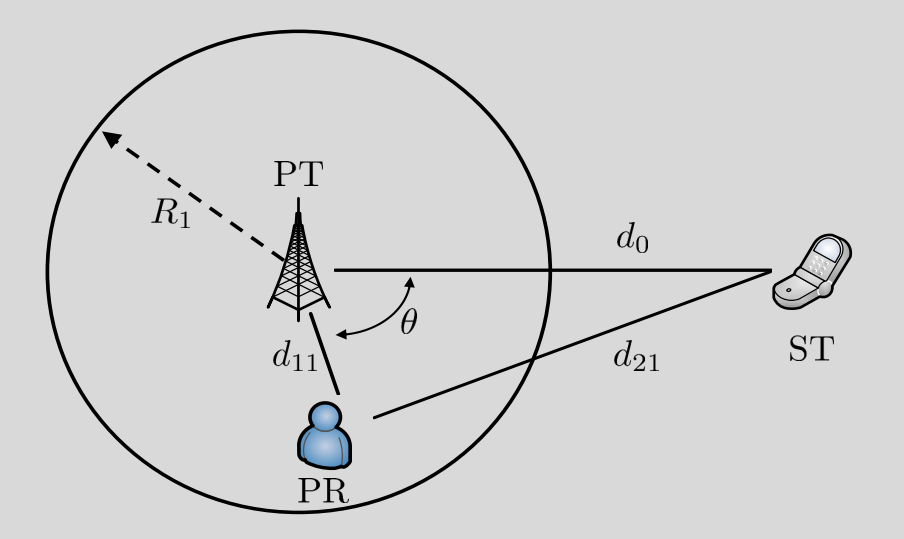

Figure 6.3: System model considering the random deployment of the users.

6.4 Interference Modeling

Location uncertainty is one of the challenges in realistic CR systems. For instance, in case the primary network is a TV network, the PRs are passive devices and their exact locations may be unknown to the SUs. On the contrary, the location of the PT can be easily obtained. Another example is a primary downlink cellular network in which the position of the BS is fixed and the mobile terminals' position information is difficult to be acquired. Given such location uncertainty, the performance degradation caused to the PUs is hard to control at the SUs. In this section, we study a way to calculate the expectation value of the interference to the PR caused by the ST considering the randomness of the PR's position. This result can be employed in Section 6.2 and Section 6.3 to design the power adaptation strategies.

The statistical CSI parameter can be obtained depending on the distance between the users. For example, using the path-loss channel model in [61, 117] and denoting the distance of the ST-PR link as d_{21}, the statistical CSI l_{21} is

$$l_{21}^{-1} = K d_{21}^{-\alpha} \tag{6.35}$$

where the path loss exponent α is within the typical range [1.6, 6] [109]. The variable K is a constant which incorporates the effect of transmit and receiver antennas and depends on the frequency band [61]. Without loss of generality, we set K = 1 in the remainder of this section. Note that the shadowing effect is currently not considered. The same relation as stated in (6.35) applies to the statistical CSI and the distances of other links, respectively.

Considering the effect of location uncertainty on the interference constraint, e.g., the left hand side of the third constraint in (6.3), the expected interference power is

$$\mathbb{E}_{\gamma, g_{22}, g_{21}} \{ P_2(\gamma, g_{22}) g_{21} | \mathcal{H}_1 \}$$

$$\begin{aligned}
&\overset{(a)}{=} \mathbb{E}_{l_{21}}\left\{\mathbb{E}_{\gamma,g_{22},g_{21}}\left\{P_2(\gamma,g_{22})g_{21}|l_{21},\mathcal{H}_1\right\}\right\}\\
&\overset{(b)}{=} \mathbb{E}_{l_{21}}\left\{\frac{\mathbb{E}_{\gamma,g_{22}}\left\{P_2(\gamma,g_{22})|\mathcal{H}_1\right\}}{l_{21}}\right\}\\
&\overset{(c)}{=} \mathbb{E}_{\gamma,g_{22}}\left\{P_2(\gamma,g_{22})|\mathcal{H}_1\right\}\mathbb{E}_{d_{21}}\left\{d_{21}^{-\alpha}\right\}
\end{aligned} \tag{6.36}$$

where, in (a), we use the definition of the conditional PDF. In (b), the mutual independence between the channel power gain g_{22} and g_{21} is applied. In (c), we incorporate (6.35) and use $K = 1$. Therefore, in order to model the interference from the ST to the PR or constrain the interference power level in the optimization problem, e.g, in the optimization problem (6.3), it is imperative for the SU to learn the distance information d_{21} and calculate the term $\mathbb{E}_{d_{21}}\left\{d_{21}^{-\alpha}\right\}$.

Figure 6.3 shows the system consisting of one PT-PR link and one ST.[3] We assume that the PR is randomly and uniformly distributed inside a circle of radius R_1 centered at the PT. The PDF of the distance of the PT-PR link d_{11} is

$$f_{d_{11}}(x) = \begin{cases} \dfrac{2x}{R_1^2}, & 0 \le x \le R_1 \\ 0, & \text{otherwise.} \end{cases} \tag{6.37}$$

The ST is assumed to be located outside the primary network with ε protection region [130], i.e., the minimum distance between the ST and the PR is ε. The distance of the PT-ST link is given by d_0 and $d_0 > R_1 + \varepsilon$ holds. The angle between the PT-PR link and the PT-ST link is denoted by θ and thus θ is uniformly distributed in $[0, 2\pi)$. The ST and the SR have the exact position knowledge of the PT but only know the distribution of the PR.

Using the law of cosines, we have

$$d_{21} = \sqrt{d_{11}^2 - 2d_{11}d_0\cos\theta + d_0^2} \tag{6.38}$$

Incorporating it into the expectation $\mathbb{E}_{d_{21}}\left\{d_{21}^{-\alpha}\right\}$, we have

$$\begin{aligned}
&\mathbb{E}_{d_{21}}\left\{d_{21}^{-\alpha}\right\}\\
&\overset{(a)}{=} \mathbb{E}_{d_{11},\theta}\left\{\left(d_{11}^2 - 2d_{11}d_0\cos\theta + d_0^2\right)^{-\alpha/2}\right\}\\
&\overset{(b)}{=} \int_0^{R_1}\int_0^{2\pi}\left(d_{11}^2 - 2d_{11}d_0\cos\theta + d_0^2\right)^{-\alpha/2}\frac{1}{2\pi}\frac{2d_{11}}{R_1^2}d\theta d(d_{11})\\
&= \frac{1}{\pi R_1^2 d_0^\alpha}\int_0^{R_1}\int_0^{\pi} 2d_{11}\left(\left(\frac{d_{11}}{d_0}\right)^2 - 2\left(\frac{d_{11}}{d_0}\right)\cos\theta + 1\right)^{-\alpha/2} d\theta d(d_{11})
\end{aligned} \tag{6.39}$$

[3] The SR is omitted herein since we focus on calculating the interference caused to the PR assuming its location ambiguity.

where, in (a), we insert (6.38). Note that since the exact knowledge of d_{21} is not available at the ST, the expectation $\mathbb{E}_{d_{21}}\left\{d_{21}^{-\alpha}\right\}$ is calculated w.r.t. the related random variables θ, d_{11}, and d_0. In (b), we incorporate the PDF of θ and d_{11}. In general, the closed-form solution of the integral (6.39) is difficult to obtain. However, for the special case that α is a even number, i.e., $\alpha = 2i$, $i \in \mathbb{N}$, it can be rewritten as

$$\begin{aligned}
&\mathbb{E}_{d_{21}}\left\{d_{21}^{-\alpha}\right\} \\
&\overset{(a)}{=} \frac{1}{\pi R_1^2 d_0^{\alpha}} \int_0^{R_1} 2d_{11}\pi \left(1-\left(\frac{d_{11}}{d_0}\right)^2\right)^{\frac{\alpha}{2}} \left(\sum_{k=0}^{\frac{\alpha}{2}-1} \frac{\left(\frac{\alpha}{2}+k-1\right)!}{(k!)^2\left(\frac{\alpha}{2}-k-1\right)!} \left(\frac{\left(\frac{d_{11}}{d_0}\right)^2}{1-\left(\frac{d_{11}}{d_0}\right)^2}\right)^k\right) d(d_{11}) \\
&\overset{(b)}{=} \frac{1}{R_1^2 d_0^{\alpha-2}} \int_0^{\left(\frac{R_1}{d_0}\right)^2} \frac{1}{(1-\xi)^{\frac{\alpha}{2}}} \sum_{k=0}^{\frac{\alpha}{2}-1} a_k^{(\alpha)} \frac{\xi^k}{(1-\xi)^k} d\xi \\
&= \frac{1}{R_1^2 d_0^{\alpha-2}} \sum_{k=0}^{\frac{\alpha}{2}-1} a_k^{(\alpha)} \int_0^{\left(\frac{R_1}{d_0}\right)^2} \frac{\xi^k}{(1-\xi)^{\left(\frac{\alpha}{2}+k\right)}} d\xi
\end{aligned} \tag{6.40}$$

where, in (a), we apply [52, Eq. 3.616.2]. In (b), we use $\xi = (d_{11}/d_0)^2$ and define

$$a_k^{(\alpha)} = \frac{\left(\frac{\alpha}{2}+k-1\right)!}{(k!)^2\left(\frac{\alpha}{2}-k-1\right)!}. \tag{6.41}$$

A closed-form expression of (6.40) for even number α can be given as follows.

- $\alpha = 2$:

$$\mathbb{E}_{d_{21}}\left\{d_{21}^{-\alpha}\right\} = \frac{1}{R_1^2} \sum_{k=0}^{\frac{\alpha}{2}-1} a_0^2 \int_0^{\left(\frac{R_1}{d_0}\right)^2} \frac{1}{(1-\xi)} d\xi = \frac{1}{R_1^2} \ln\left(\frac{d_0^2}{d_0^2 - R_1^2}\right). \tag{6.42}$$

- $\alpha = 4, 6, 8, \ldots$: we define

$$I(k;\alpha) = \int_0^{\left(\frac{R_1}{d_0}\right)^2} \xi^k (1-\xi)^{-\left(\frac{\alpha}{2}+k\right)} d\xi \tag{6.43}$$

and thus Eq. (6.40) is reformulated as

$$\mathbb{E}_{d_{21}}\left\{d_{21}^{-\alpha}\right\} = \frac{1}{R_1^2 d_0^{\alpha-2}} \sum_{k=0}^{\frac{\alpha}{2}-1} a_k^{(\alpha)} I(k;\alpha). \tag{6.44}$$

Using integration by parts, we find $I(k;\alpha)$, $\forall\, k > 0$, can be calculated recursively

$$I(k;\alpha) = \frac{\xi^k}{1-\frac{\alpha}{2}-k} (1-\xi)^{1-\frac{\alpha}{2}-k} \Bigg|_0^{\left(\frac{R_1}{d_0}\right)^2}$$

$$
\begin{aligned}
&- \frac{1}{1-\frac{\alpha}{2}-k}\int_0^{\left(\frac{R_1}{d_0}\right)^2} \xi^{k-1}(1-\xi)^{-\left(\frac{\alpha}{2}+k-1\right)}d\xi \\
&= \frac{\xi^k}{1-\frac{\alpha}{2}-k}(1-\xi)^{1-\frac{\alpha}{2}-k}\Bigg|_0^{\left(\frac{R_1}{d_0}\right)^2} - I(k-1;\alpha) \qquad (6.45)
\end{aligned}
$$

with

$$
I(0;\alpha) = \frac{(1-\xi)^{1-\alpha}}{1-\frac{\alpha}{2}}. \qquad (6.46)
$$

For example, for the path loss exponent $\alpha = 4$, we have

$$
\mathbb{E}_{d_{21}}\left\{d_{21}^{-\alpha}\right\} = \frac{\left(a_0^{(4)}I(0;4) + a_1^{(4)}I(1;4)\right)}{R_1^2 d_0^2} = \frac{1}{(R_1^2 - d_0^2)^2}. \qquad (6.47)
$$

So far, we have discussed the closed-form expression of $\mathbb{E}_{d_{21}}\left\{d_{21}^{-\alpha}\right\}$ for positive and even numbers of α. In a realistic scenario, the path loss exponent α typically varies from 1.6 to 6. Therefore, it is also desirable to consider the cases where α is any fraction within this region. However, a closed-form expression for this general case is difficult to derive. We resort to two alternatives. On the one hand, some existing numerical integration methods, e.g., trapezoidal numerical integration, can be applied to (6.39). The details are omitted here. On the other hand, we can derive an upper bound of $\mathbb{E}_{d_{21}}\left\{d_{21}^{-\alpha}\right\}$ denoted as D_U. The purpose to calculate its upper bound is reasoned as follows. Replacing $\mathbb{E}_{d_{21}}\left\{d_{21}^{-\alpha}\right\}$ by its upper bound D_U in (6.36), we have

$$
\begin{aligned}
&\mathbb{E}_{\gamma,g_{22}}\left\{P_2(\gamma,g_{22})|\mathcal{H}_1\right\}\mathbb{E}_{d_{21}}\left\{d_{21}^{-\alpha}\right\} \le \mathbb{E}_{\gamma,g_{22}}\left\{P_2(\gamma,g_{22})|\mathcal{H}_1\right\}D_U \le P_I \\
&\Rightarrow \mathbb{E}_{\gamma,g_{22}}\left\{P_2(\gamma,g_{22})|\mathcal{H}_1\right\} \le P_I/D_U \le P_I/\mathbb{E}_{d_{21}}\left\{d_{21}^{-\alpha}\right\} \qquad (6.48)
\end{aligned}
$$

where P_I is the interference power limit at the PR. Eq. (6.48) indicates that the power constraint is bounded by the lower bound of the original power limit, thus a stricter power constraint is imposed. Incorporating this into the optimization problem, we obtain a suboptimal power allocation solution that strictly guarantees the interference caused to the PUs.

In order to establish D_U, we consider the system model shown in Figure 6.4. Compared to Figure 6.3, apart from the primary network centered at the PT with the radius R_1, we additionally consider two networks centered at the ST with radii $\sqrt{d_0^2 - R_1^2}$ and $d_0 - R_1$ which are drawn with dash-dotted and dashed lines, respectively. It is easy to verify that $\sqrt{d_0^2 - R_1^2} \ge d_0 - R_1$ holds under the previous assumption $d_0 > R_1 + \varepsilon$. To calculate the upper bound D_U, we assume the PR is only uniformly distributed in the shadowed region. Since the PR in this region is in general closer to the ST compared to the primal assumption that the PR is uniformly deployed in the circle centered at the PT, the expected value $\mathbb{E}_{d_{21}}\left\{d_{21}^{-\alpha}\right\}$ is upper-bounded for positive α.

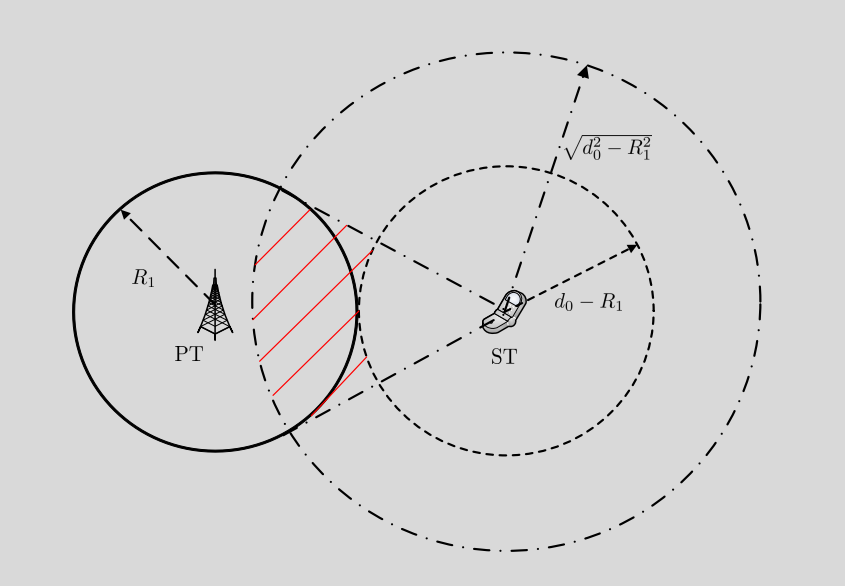

Figure 6.4: System model considering the random deployment of the users. The solid circle denotes the primary network. The dash-dotted and dashed circles are centered at the ST with radii $\sqrt{d_0^2 - R_1^2}$ and $d_0 - R_1$, respectively. The PR is uniformly distributed in the shadowed region.

Under the aforementioned assumptions, the PDF of the distance d_{21} of the ST-PR link is

$$f_{d_{21}}(x) = \begin{cases} \dfrac{2x}{d_0^2 - R_1^2 - (d_0 - R_1)^2}, & d_0 - R_1 \leq x \leq \sqrt{d_0^2 - R_1^2} \\ 0, & \text{otherwise.} \end{cases} \tag{6.49}$$

Thus, the upper bound of $\mathbb{E}_{d_{21}}\left\{d_{21}^{-\alpha}\right\}$ is computed as

$$\begin{aligned} D_U &= \int_{d_0 - R_1}^{\sqrt{d_0^2 - R_1^2}} d_{21}^{-\alpha} f_{d_{21}}(d_{21}) d(d_{21}) \\ &= \begin{cases} \dfrac{2}{(\alpha - 2)\left(d_0^2 - R_1^2 - (d_0 - R_1)^2\right)} \left(\dfrac{1}{(d_0 - R_1)^{\alpha - 2}} - \dfrac{1}{\left(\sqrt{d_0^2 - R_1^2}\right)^{\alpha - 2}} \right), & \alpha \neq 2 \\ \dfrac{2}{\left(d_0^2 - R_1^2 - (d_0 - R_1)^2\right)} \left(\ln\left(\left(\sqrt{d_0^2 - R_1^2}\right)\right) - \ln\left(d_0 - R_1\right) \right), & \text{otherwise.} \end{cases} \end{aligned}$$

For realistic scenarios, we only need to use this analytical result for $\alpha \in [1.6, 6]$ and α is not necessarily an even number. If α is the even number, e.g., $\alpha = 2$, we use the derived exact expression in (6.42).

The accuracy of interference modeling is verified in Figure 6.5. Assuming the PT is located at the origin of the cartesian coordinate system. A ST is positioned at $(2, 2)$.

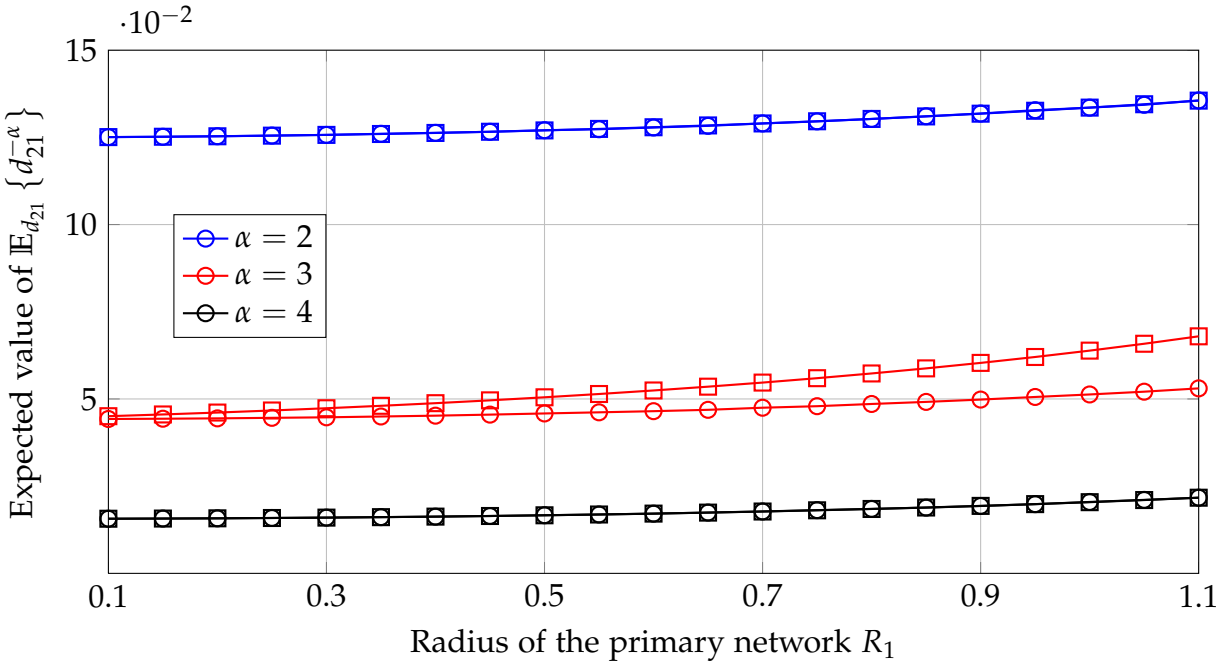

Figure 6.5: Expected value of $\mathbb{E}_{d_{21}}\{d_{21}^{-\alpha}\}$ vs. the radius of the primary network R_1 under different path loss exponents. Solid lines with squares: analytical results (exact value for even number α and an upper bound for a general value of α). Solid lines with circles: numerical results.

The primary network is a circle centered at the PT and R_1 is the radius varying from 0.1 to 1.1. The exemplified path loss exponent α is chosen to be 2, 3, and 4. The results show that the analytical result matches the numerical evaluation well for even numbers of α. On the other hand, the upper bound of $\mathbb{E}_{d_{21}}\{d_{21}^{-\alpha}\}$ is more accurate when the radius R_1 is smaller, i.e., the ST is far away from the primary network.

Figure 6.6 shows the achievable rate of the secondary link vs. the transmit power constraint P_S. The random deployment of the PR is assumed with the path loss exponent $\alpha = 2$ (free space). The detailed explanations of the plots are given below the figure. Similar to Figure 6.2, the near-optimality of the suboptimal algorithm for the hybrid paradigm is also verified. Moreover, the superiority of the power adaptation strategies in the hybrid paradigm is validated against the standard underlay and the interweave paradigm.

6.5 Summary

In this chapter, we have verified the effectiveness of a hybrid paradigm, i.e., a combination of the standard interweave and underlay paradigms. Specifically, we allowed the SU to exploit the spectrum sensing information at the ST to adapt its transmission parameters, e.g., the power level. The effect of location uncertainty is also considered in modeling the interference caused to the primary system. Several observations are obtained: On the one hand, compared to the power allocation in the interweave

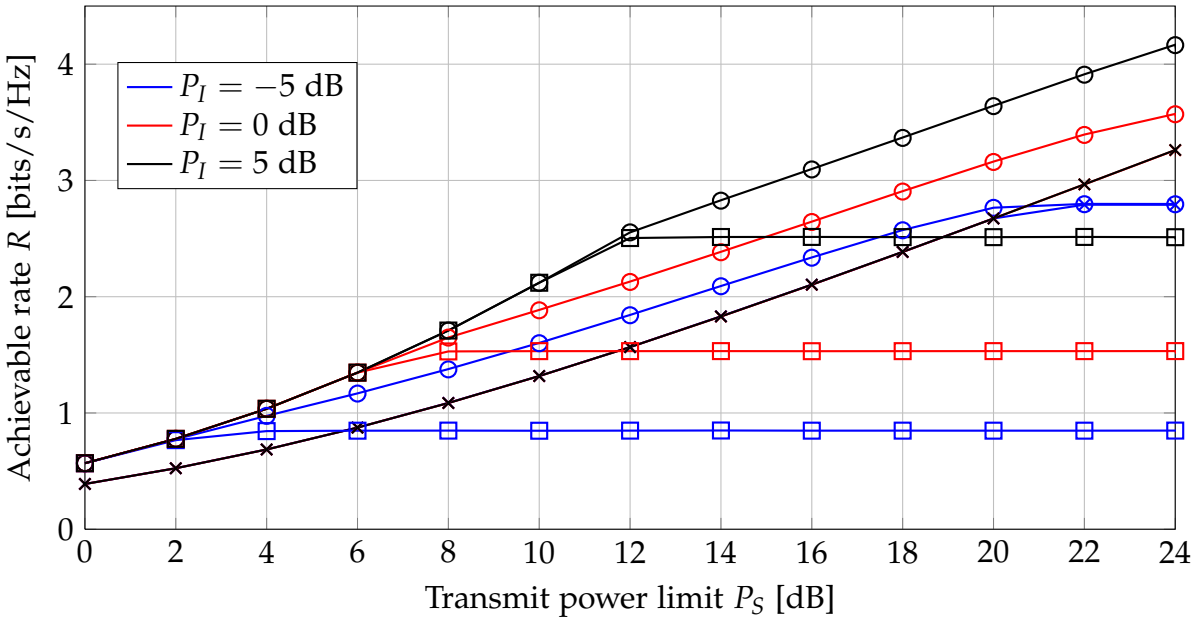

Figure 6.6: Achievable rate of the secondary link vs. the transmit power limit P_S with different interference power limits P_I. The random deployment of the PR is assumed. Solid lines: optimal algorithm for the hybrid paradigm based on hard-sensing decisions; circles: suboptimal algorithm for hybrid paradigm based on hard-sensing decisions; solid lines marked with crossings: suboptimal algorithm for the interweave paradigm; solid lines marked with squares: suboptimal algorithm for the underlay paradigm.

paradigm, the performance gain achieved by the proposed strategy in the hybrid paradigm is mainly due to the additional transmission during the period when the presence of the primary user is detected. On the other hand, given the reliable spectrum sensing results and comparably short sensing duration, the proposed scheme outperforms the underlay paradigm due to the additional utilization of sensing observations.

Overall, we have studied the joint system design of the sensing and transmission functionals. During the last few decades, research focusing on some specifical functionals for the secondary system has been very active, e.g., on spectrum sensing, access protocol, or transceiver design. Therefore, we envision large potentials in the joint consideration of these functionals in a CR system design to finally achieve "real cognition".

Chapter 7

Conclusions and Outlook

Recognized as a promising technology, cognitive radio allows for the coexistence of both licensed and unlicensed devices. It builds on the environment-aware dynamic spectrum access aiming at increasing spectral efficiency and restraining the performance degradation to the incumbent devices. This demands for a proper interference management for secondary systems, including the identification of the PUs' activity and adaptation of the transmission parameters accordingly. However, the design of strategies at the SUs is challenging due to limited cooperation between the PUs and the SUs. Driven by this fact, the focus of this thesis is a comprehensive and realistic study on interference management strategies and the characterization of the achievable performance of the secondary link in secondary systems.

Spectrum sensing aims at detecting the presence or absence of the PUs. Such sensing decision can be exploited for interference avoidance in opportunistic transmission, or the adjustment of transmission parameters in spectrum sharing. The main challenge encountered in spectrum sensing is the high requirement on sensitivity, reliability, and agility, especially in the low SNR region with environment uncertainties, e.g., fading channels, noise uncertainty, and the hidden primary user problem. Cooperative spectrum sensing is considered in order to address the hidden primary user problem by exploiting the DoFs offered by multiple SUs. Mathematically, such problems subject to environment uncertainties are formulated as binary hypothesis problems with unknown parameters which can be effectively solved by the GLRT principle. In Chapter 3, we applied the GLRT framework and shown how to efficiently utilize the limited a priori knowledge at the SUs to combat the environment uncertainties. Specifically, on the one hand, the usage of rank information to extract the structure of the primary signal space was elaborated. Based on the estimation of the signal structure, subspace-based cooperative sensing methods are proposed which differentiate between binary hypotheses using the distinct correlation properties of the received signals. The quantitative comparison with other conventional cooperative sensing methods has revealed the effectiveness of using the rank information and the robustness to the uncertainties of the signal space and noise variances. On the other hand, we considered the effect of the fading sensing and reporting channels on cooperative spectrum sensing. Due to practical challenges, the SUs only have partial CSI of the channels and do not know the exact structure of the primary signal space. We derived GLRT-based sensing algorithms under two circumstances, i.e., when the noise variances of the sensing channels are unknown or known. The resulting methods counter the performance degradation introduced by the fading effects and the noise uncertainty.

Interference mitigation techniques are inevitably required in spectrum sharing systems. For example, the concurrent operation of the PUs and the SUs occurs in the underlay and the overlay paradigm, as well as in case of a missed detection in the interweave paradigm. Contrary to conventional interference networks consisting of users with equal priority to access the resources, the unique challenge in DSA is the limited preliminary knowledge related to primary systems available at the SUs. Based on this limited information, the SUs need to optimize their transceiver strategies, such as transmit power, bandwidth, or precoders and equalizers. The goal is to achieve the tradeoff between improved secondary network throughput and, most critically, constrain the performance loss of the primary transmission. Moreover, given the developed strategies, it is also desirable to quantitatively characterize the achievable performance of the SUs at the expense of the performance degradation to the PUs. Such an analysis can be used for a performance assessment or facilitate the decision in selecting system parameters. Hence, we addressed these issues in Chapter 4 and Chapter 5. In particular, for a single-antenna spectrum sharing system, we studied power allocation strategies for the SUs subject to different QoS constraints on the primary link: a traditional interference temperature constraint and an outage probability constraint. Not only the optimal and low-complexity near-optimal power allocation strategies are developed, but also the achievable performance of the system is approximately evaluated in closed form. Additionally, for sophisticated multi-antenna infrastructure-based spectrum sharing networks, we aimed at developing the robust transceiver filters at the SUs subject to the interference temperature constraint under the consideration of imperfect CSI. Two commonly-assumed CSI error models are used: a bounded CSI error model and a stochastic CSI error model. The difficulty in solving such problems lies in their non-convexity. We derived several efficient algorithms by applying different approaches. For example, we proposed the alternating method in which the non-convex primal problem is decomposed and reformulated into convex subproblems, the constrained gradient projection method which is in general more efficient than the methods depending on convex optimization solvers, and the method based on uplink–downlink duality in which the favorable property of the dual uplink problem can be exploited to solve the primal downlink problem. Through a performance evaluation and comparison, we verified that the robust transceiver design is able to strictly limit the performance degradation of the PUs in spite of the imperfect CSI at the SUs, which is a critical criterion for the design of the secondary transmission. Furthermore, the advantage of exploiting the constrained gradient projection method, especially by combining this method with uplink–downlink duality, is validated to solve the downlink optimization problem in terms of faster convergence rate and better target value.

So far, spectrum sensing and interference mitigation strategies were investigated only in a single cognitive paradigm, e.g., either in the interweave paradigm or the underlay paradigm, respectively. In Chapter 6, we studied a hybrid paradigm combining the interweave and the underlay paradigm. Specifically, we extended the design of power allocation strategies in Chapter 4 to the scenario allowing the ST to exploit the spectrum sensing information to adapt the power level. The effect of location uncer-

tainties is also considered in modeling the interference caused to the primary system. Both soft-decision and hard-decision sensing results are explored. We found that compared to the power allocation in the interweave paradigm, the proposed strategy in the hybrid paradigm achieves better performance due to transmission upon the detection of the PUs' activity, whereas it outperforms the underlay paradigm due to the additional utilization of sensing observations given reliable sensing results and comparably short sensing durations. Therefore, we have exemplified the essential "cognitive" property that acquiring the useful information from the environment and utilizing such information improves the spectrum efficiency.

Within this thesis, we have conducted research on advanced interference management techniques in CR systems by considering realistic challenges. For the future work, we envision the following topics.

- *Optimization of spectrum sensing parameters*: In the proposed sensing methods, the spectrum sensing parameters, e.g., the sensing observation length and the number of cooperative users, are fixed. It would be interesting to dynamically optimize such parameters, e.g., to optimize the sensing and transmission length jointly in one frame to achieve the tradeoff between the sensing effort and the resulting throughput gain. Furthermore, due to the difficulty in deriving closed-form relations between the probability of detection and the false alarm rate, the decision threshold is in general hard to optimize. Therefore, an exact or an asymptotic performance analysis of certain sensing methods, e.g., similar to [10,136], is imperative for the selection of the threshold.

- *Extension of power control strategies to the multiuser scenario*: Based on our research on the power allocation strategies in a spectrum sharing system consisting of a single primary and a single secondary link, we are interested in answering the question of how the current result scales with a larger number of users in the secondary network. Moreover, cognitive transmission is potentially suitable for short-range transmission due to its low power emission property [54]. Therefore, the extension of the present work to such short-range transmission is also of interest, including how multiple SUs with different QoS requirements exploit the dynamic power allocation to share the available spectrum opportunities in space and time.

- *Study on the exploitation of general learning functionalities*: Apart from spectrum sensing results, large performance gains are expected for the SUs by using the results from more general learning functionalities, e.g., the traffic-intensity model of the PUs and transmission adaptation patterns used by the PTs. Recent research on certain functionalities has been very active, for example, spectrum analysis, spectrum decision, and spectrum sharing. We envision large potentials in the joint design of these functionalities for CR systems.

Appendix A

Mathematical Derivations

A.1 Solutions of $[\boldsymbol{\mu}_i]_k$ and $[\boldsymbol{\Sigma}_i]_{k,k'}$

The kth element of $\boldsymbol{\mu}_i$ is

$$\begin{aligned}
[\boldsymbol{\mu}_i]_k = \ \mathbb{E}\left(y_k|\mathcal{H}_i\right) &= \int_{x_k,g_k,y_k} y_k p\left(y_k|x_k,g_k,\mathcal{H}_i\right) p\left(x_k,g_k|\mathcal{H}_i\right) dy_k dg_k dx_k \\
&= \int_{x_k} p\left(x_k|\mathcal{H}_i\right) \int_{g_k} p\left(g_k\right) \int_{y_k} y_k p\left(y_k|x_k,g_k,\mathcal{H}_i\right) dy_k dg_k dx_k \\
&= \int_{x_k} p\left(x_k|\mathcal{H}_i\right) \int_{g_k} g_k x_k p\left(g_k\right) dg_k dx_k \\
&= \bar{g}_k \bar{x}_{k,i}. \qquad \text{(A.1)}
\end{aligned}$$

where $\bar{g}_k = \mathbb{E}\left(g_k\right)$ and $\bar{x}_{k,i} = \mathbb{E}\left(x_k|\mathcal{H}_i\right)$.

The kth row and $k^{'}$th column of the covariance matrix is

$$\begin{aligned}
[\boldsymbol{\Sigma}_i]_{k,k'} &= \left[\mathbb{E}\left\{\mathbf{y}\mathbf{y}^H|\mathcal{H}_i\right\} - \boldsymbol{\mu}_i\boldsymbol{\mu}_i^H\right]\Big]_{k,k'} \\
&= \int_{y_k,y_{k'}} \left(y_k - [\boldsymbol{\mu}_i]_k\right)\left(y_{k'} - [\boldsymbol{\mu}_i]_{k'}\right)^* p\left(y_k,y_{k'}|\mathcal{H}_i\right) dy_k dy_{k'} \\
&\overset{(a)}{=} \int_{y_k,y_{k'}} \left(y_k - [\boldsymbol{\mu}_i]_k\right)\left(y_{k'} - [\boldsymbol{\mu}_i]_{k'}\right)^* \int_{x_k,x_{k'},g_k,g_{k'}} p\left(y_k,y_{k'}|x_k,x_{k'},g_k,g_{k'},\mathcal{H}_i\right) \\
&\quad p\left(x_k,x_{k'},g_k,g_{k'}|\mathcal{H}_i\right) dx_k dx_{k'} dg_k dg_{k'} dy_k dy_{k'} \\
&\overset{(b)}{=} \int_{x_k,x_{k'}} p\left(x_k,x_{k'}|\mathcal{H}_i\right) \int_{g_k,g_{k'}} p\left(g_k,g_{k'}\right) \\
&\quad \int_{y_k,y_{k'}} \left(y_k - g_k x_k + g_k x_k - [\boldsymbol{\mu}_i]_k\right)\left(y_{k'} - g_{k'}x_{k'} + g_{k'}x_{k'} - [\boldsymbol{\mu}_i]_{k'}\right)^* \\
&\quad p\left(y_k,y_{k'}|x_k,x_{k'},g_k,g_{k'},\mathcal{H}_i\right) dy_k dy_{k'} dg_k dg_{k'} dx_k dx_{k'} \\
&\overset{(c)}{=} \int_{x_k,x_{k'}} p\left(x_k,x_{k'}|\mathcal{H}_i\right) \int_{g_k,g_{k'}} p\left(g_k,g_{k'}\right) \\
&\quad \int_{y_k,y_{k'}} \left(v_k v_{k'}^* + \left(g_k x_k - [\boldsymbol{\mu}_i]_k\right)\left(g_{k'}x_{k'} - [\boldsymbol{\mu}_i]_{k'}\right)^*\right) \\
&\quad p\left(y_k,y_{k'}|x_k,x_{k'},g_k,g_{k'},\mathcal{H}_i\right) dy_k dy_{k'} dg_k dg_{k'} dx_k dx_{k'} \\
&= \sigma_v^2 \delta_{k,k'} + \mathbb{E}\left\{x_k x_{k'}^*\right\}\mathbb{E}\left\{g_k g_{k'}^*\right\} - [\boldsymbol{\mu}_i]_k\left[\boldsymbol{\mu}_i\right]_{k'}^* \qquad \text{(A.2)}
\end{aligned}$$

where step (a) uses the Bayesian rule to expand the conditional probability $p(y_k, y_{k'}|\mathcal{H}_i)$ as a function of $x_k, x_{k'}, g_k$, and $g_{k'}$. Step (b) expresses $p(x_k, x_{k'}, g_k, g_{k'}|\mathcal{H}_i)$ as the product of $p(x_k, x_{k'}|\mathcal{H}_i)$ and $p(g_k, g_{k'})$ by exploiting the mutual independency between the received signals and the reporting channels. The next step (c) simplifies the third integral in the last step by considering that $y_k - g_k x_k = v_k$ for all k and removes the items including the first order of v_k and $v_k^{'}$ because theirs means are both equal to zero.

A.2 Analytical Form of $Q_{avg}(v)$ and $Q_{avg}^{'}(v)$

The key to obtain closed-form expression of $Q_{avg}(v)$ and $Q_{avg}^{'}(v)$ is to derive

$$Q_1(T) = \int_0^T f(t)dt, \tag{A.3}$$

$$Q_2(T) = \int_0^T tf(t)dt. \tag{A.4}$$

Recalling that $t = (\sigma_s^2 + P_1 g_{12})/g_{22}$, the variables g_{22} and $P_1 g_{12}$ follow an exponential distribution with the mean l_{22}^{-1} and P_1/l_{12}, respectively, we follow a similar approach as in [12, (48)] and derive the probability density distribution function of t as

$$f(t) = \begin{cases} \dfrac{l_{22}l_y^2\sigma_s^2 + l_{22}l_y + \frac{l_{22}^2 l_y \sigma_s^2}{t}}{(l_{22} + l_y t)^2} e^{\left(\frac{-l_{22}\sigma_s^2}{t}\right)} & t \geq 0 \\ 0 & t < 0 \end{cases} \tag{A.5}$$

with $l_y = l_{21}/P_1$. Taking (A.5) into (A.3), we have

$$\begin{aligned} Q_1(T) &\overset{(a)}{=} a \int_{\tilde{\gamma}}^{\infty} \frac{z + a + 1}{(z+a)^2} e^{-z} dz \\ &= a \left(\int_{\tilde{\gamma}}^{\infty} \frac{1}{(z+a)} e^{-z} dz + \int_{\tilde{\gamma}}^{\infty} \frac{1}{(z+a)^2} e^{-z} dz \right) \\ &\overset{(b)}{=} a \left(e^a E_1(\tilde{\gamma} + a) + \frac{1}{\tilde{\gamma} + a} e^{-\tilde{\gamma}} - e^a E_1(\tilde{\gamma} + a) \right) \\ &= \frac{a}{\tilde{\gamma} + a} e^{-\tilde{\gamma}}. \end{aligned}$$

where in step (a) we use the following replacements to simplify the integral:

$$a = \frac{l_{12}}{P_1}\sigma_s^2, \tilde{\gamma} = \frac{l_{22}}{T}\sigma_s^2, z = \frac{l_{22}}{t}\sigma_s^2.$$

In step (b), we use the definition of the exponential integral [4]

$$E_1(x) = \int_x^{\infty} \frac{e^{-t}}{t} dt, \quad \mathcal{R}(x) > 0.$$

Considering $Q_2(T)$, we take (A.5) into its expression and get

$$\begin{aligned} Q_2(T) &\overset{(a)}{=} a\int_{\gamma}^{\infty}\frac{z+g+1}{z(z+g)^2}e^{-z}dz \\ &= a\int_{\gamma}^{\infty}\left(\left(\frac{1}{z}-\frac{1}{z+g}\right)\frac{1}{g}+\frac{1}{g^2}\left(-\frac{1}{z+g}-\frac{g}{(z+g)^2}+\frac{1}{z}\right)\right)e^{-z}dz \\ &= \left(\frac{g}{a}+\frac{g}{a^2}\right)E_1(\gamma)-\frac{g}{a^2}e^{a}E_1(\gamma+a)-\frac{g}{a(\gamma+a)}e^{-\gamma} \end{aligned}$$

where in step (a), we use

$$g=\frac{l_{12}l_{22}}{P_1}\sigma_s^4.$$

Consequently, $Q_{avg}(v)$ and $Q_{avg}^{'}(v)$ in (4.18) and (4.20) can be represented as

$$Q_{avg}(v)=(P_S-v)Q_1(v-P_S)+vQ_1(v)-(Q_2(v)-Q_2(v-P_S)) \tag{A.6}$$

$$Q_{avg}^{'}(v)=Q_1(v)-Q_1(v-P_S). \tag{A.7}$$

A.3 Proof of Proposition 4.2.2

We first define the function $r(T,\alpha,\beta)$ with $\alpha+\beta/T>0$ which will be used later as

$$r(T,\alpha,\beta)=\int_0^T\ln\left(\alpha+\frac{\beta}{t}\right)f(t)dt. \tag{A.8}$$

It is easy to see that $r(T,\alpha,\beta)=0$ for $T<0$. In the following, we express the achievable rate as the function of $r(T,\alpha,\beta)$. Specifically, integrating the optimal power control solution (4.10) into the objective function of (4.8), we obtain

$$R=\int_0^{v^\star-P_S}\ln\left(1+\frac{P_S}{t}\right)f(t)dt+\int_{v^\star-P_S}^{v^\star}\ln\left(\frac{v^\star}{t}\right)f(t)dt. \tag{A.9}$$

Using (A.9), after some simple mathematical computations, we can rewrite (A.9) as (4.22) for $P_S\le P_I l_{21}$. Similarly, the performance is given in (4.23) for $P_S>P_I l_{21}$.

Therefore, the key point to derive Proposition 4.2.2 is to calculate the closed-form expression of $r(T,\alpha,\beta)$, $T\ge 0$. Specifically, we take the expression of $f(t)$ in (A.5) into (A.9) and obtain

$$\begin{aligned} r(T,\alpha,\beta) &= \int_0^T\ln\left(\alpha+\frac{\beta}{t}\right)\frac{l_{22}(l_{12}/P_1)^2\sigma_S^2+l_{22}l_{12}/P_1+\frac{l_{22}^2l_{12}\sigma_S^2/P_1}{t}}{(l_{22}+l_{12}t/P_1)^2}e^{\left(\frac{-l_{22}\sigma_S^2}{t}\right)}dt \\ &\overset{(a)}{=} \int_{\gamma}^{\infty}b\ln(\alpha+x)\frac{x+h+1/g}{(x+h)^2}e^{-gx}dx \\ &= b\int_{\gamma}^{\infty}\ln(\alpha+x)\left(\frac{1}{x+h}+\frac{1/g}{(x+h)^2}\right)e^{-gx}dx \end{aligned} \tag{A.10}$$

where, in (a), we use the following notation for brevity

$$\gamma = \frac{\beta}{T},\ x = \frac{\beta}{t},\ b = \frac{l_{12}\sigma_S^2}{P_1},\ g = \frac{l_{22}\sigma_S^2}{\beta},\ h = \frac{l_{12}\beta}{l_{22}P_1} \tag{A.11}$$

Similar to [12, Eq. (54)], we consider the following integration

$$\begin{aligned}
&\int_\gamma^\infty \ln(\alpha+x)\frac{1}{x+h}e^{-gx}dx \\
&\overset{(a)}{=}\frac{1}{g}\left(\ln(\alpha+\gamma)\frac{1}{\gamma+h}e^{-g\gamma} - \int_\gamma^\infty \ln(\alpha+x)\frac{1}{(x+h)^2}e^{-gx}dx + \int_\gamma^\infty \frac{1}{(x+h)(x+\alpha)}e^{-gx}dx\right)
\end{aligned}$$

where we use integration by parts in (a). Hence, equation (A.10) is equal to

$$r(T,\alpha,\beta) = \frac{b}{g}\left(\ln(\alpha+\gamma)\frac{1}{\gamma+h}e^{-g\gamma} + \int_\gamma^\infty \frac{1}{(x+h)(x+\alpha)}e^{-gx}dx\right). \tag{A.12}$$

Different from [12, Appendix A], we calculate $r(T,\alpha,\beta)$ in closed form by distinguishing between two cases. Particularly, For the cases $h \neq \alpha$ and $h = \alpha$, we derive from (A.12) and have the expression of $r(T,\alpha,\beta)$ given in

$$\begin{aligned}
r(T,\alpha,\beta) =& \frac{b}{g}\left(\ln(\alpha+\gamma)\frac{1}{\gamma+h}e^{-g\gamma} + \int_\gamma^\infty \left(\frac{1}{x+h} - \frac{1}{x+\alpha}\right)\frac{1}{\alpha-h}e^{-gx}dx\right) \\
=& \frac{b}{g}\left(\ln(\alpha+\gamma)\frac{1}{\gamma+h}e^{-g\gamma} + \frac{1}{\alpha-h}\left(e^{gh}E_1(g(h+\gamma)) - e^{g\alpha}E_1(g(\alpha+\gamma))\right)\right), \\
& h \neq \alpha
\end{aligned} \tag{A.13}$$

$$\begin{aligned}
r(T,\alpha,\beta) =& \frac{b}{g}\left(\ln(\alpha+\gamma)\frac{1}{\gamma+h}e^{-g\gamma} + \int_\gamma^\infty \frac{1}{(x+\alpha)^2}e^{-gx}dx\right) \\
=& \frac{b}{g}\left(\ln(\alpha+\gamma)\frac{1}{\gamma+\alpha}e^{-g\gamma} + e^{-g\gamma}\frac{1}{\gamma+\alpha} - ge^{g\alpha}E_1(g(\gamma+\alpha))\right), \quad h = \alpha.
\end{aligned} \tag{A.14}$$

A.4 Monotonicity of $G(x)$

Calculating the derivative $G(x)$ over x results in $G^{'}(x) = e^x E_1(x) - 1/x$. Now consider the following inequality holds:

$$\frac{1}{x+1} \overset{(a)}{<} e^x E_1(x) \overset{(b)}{<} \ln\left(1+\frac{1}{x}\right) \overset{(c)}{<} \frac{1}{x} \quad \forall x > 0 \tag{A.15}$$

where we apply [4, Eq. 5.1.20] in (a) and (b). In (c), we use $\ln(1+x) < x$, $x > 0$. Finally, we conclude that $G^{'}(x)$ is negative and $G(x)$ is strictly monotonically decreasing w.r.t. x.

A.5 Monotonicity of $H(x)$

Calculating the derivative of $H(x)$ over x and rearranging the resulting terms, we obtain

$$\begin{aligned}\frac{dH(x)}{dx} &= b_1\Bigg(e^{a_1+b_1P_Sx}E_1(a_1+b_1P_Sx)\\ &\qquad + b_1P_Sx\left(e^{a_1+b_1P_Sx}E_1(a_1+b_1P_Sx) - \frac{1}{a_1+b_1P_Sx}\right)\Bigg)\\ &\overset{(a)}{=} b_1\left((y+1)\left(e^{y}E_1(y) - \frac{1}{y+1}\right) - a_1\left(e^{y}E_1(y) - \frac{1}{y}\right)\right) \end{aligned} \tag{A.16}$$

where, in (a), we use $y = a_1 + b_1P_Sx$ and rearrange the terms. Applying (A.15), the derivative in (A.16) is positive. Hence, we conclude that $H(x)$ is a monotonically increasing function of x.

Moreover, applying (A.15) to $H(x)$ yields $b_1x/(a_1+b_1P_Sx+1) < H(x) < b_1x/(a_1+b_1P_Sx)$. Using this inequality and the positivity of a_1, b_1, and P_S, it is straightforward to see that both the lower and upper bounds of $H(x)$ converge to zero when x approaches zero, thus we have $\lim_{x\to 0} H(x) = 0$.

A.6 Proof of Proposition 4.2.3

If $P_S > P_Il_{21}$, the finite and positive $v_2^\star$ is chosen to satisfy $\mathbb{E}_{g_{22}}\{\tilde{P}_2^\star(g_{22})\} = P_Il_{21}$. We first reformulate $\mathbb{E}_{g_{22}}\{\tilde{P}_2^\star(g_{22})\}$ as a function of $v_2^\star$.

For the case $v_2^\star > P_S$, integrating (4.60) into $\mathbb{E}_{g_{22}}\{\tilde{P}_2^\star(g_{22})\}$ yields

$$\begin{aligned}\mathbb{E}_{g_{22}}\{\tilde{P}_2^\star(g_{22})\} &= \int_{\frac{a_1+1}{b_1v_2^\star}}^{\frac{a_1+1}{b_1\left(v_2^\star-P_S\right)}} \tilde{p}^\star(x)f_{g_{22}}(x)dx + P_S\int_{\frac{a_1+1}{b_1\left(v_2^\star-P_S\right)}}^{\infty} f_{g_{22}}(x)dx\\ &= P_Se^{-\frac{(a_1+1)l_{22}}{b_1\left(v_2^\star-P_S\right)}} + v_2^\star\left(e^{-\frac{(a_1+1)l_{22}}{b_1v_2^\star}} - e^{-\frac{(a_1+1)l_{22}}{b_1\left(v_2^\star-P_S\right)}}\right)\\ &\quad - \frac{(a_1+1)l_{22}}{b_1}\left(E_1\left(\frac{(a_1+1)l_{22}}{b_1v_2^\star}\right) - E_1\left(\frac{(a_1+1)l_{22}}{b_1\left(v_2^\star-P_S\right)}\right)\right)\\ &\overset{(a)}{=} v_2^\star e^{-\frac{\lambda_b}{v_2^\star}} - \lambda_bE_1\left(\frac{\lambda_b}{v_2^\star}\right)\\ &\quad - \left((v_2^\star-P_S)\,e^{-\frac{\lambda_b}{v_2^\star-P_S}} - \lambda_bE_1\left(\frac{\lambda_b}{v_2^\star-P_S}\right)\right)\end{aligned} \tag{A.17}$$

where in (a), we introduce

$$\lambda_b = \frac{(a_1+1)l_{22}}{b_1}. \tag{A.18}$$

Similarly, for the case $0 < v_2^\star \leq P_S$, we integrate (4.61) into $\mathbb{E}_{g_{22}}\{\tilde{P}_2^\star(g_{22})\}$ and have

$$\mathbb{E}_{g_{22}}\{\tilde{P}_2^\star(g_{22})\} = \int_{\frac{a_1+1}{b_1 v_2^\star}}^{\infty} \tilde{p}^\star(x) f_{g_{22}}(x) dx = v_2^\star e^{-\frac{\lambda_b}{v_2^\star}} - \lambda_b E_1\left(\frac{\lambda_b}{v_2^\star}\right). \tag{A.19}$$

Using the definition of $F_2(x)$ in (4.63), we combine the expressions in (A.17) and (A.19) in the general form $\mathbb{E}_{g_{22}}\{\tilde{P}_2^\star(g_{22})\} = F_2(v_2^\star)$.

Therefore, in order to choose $v_2^\star$ satisfying $\mathbb{E}_{g_{22}}\{\tilde{P}_2^\star(g_{22})\} = P_I l_{21}$, we need to choose $v_2^\star$ as the root $x^\star$ of $F_2(x) = P_I l_{21}$. Note that x is used instead of v_2 here.

Applying the Newton method to root-searching for $F_2(x) = P_I l_{21}$ results in

$$x^{(n+1)} = x^{(n)} - \frac{F_2\left(x^{(n)}\right) - P_I l_{21}}{F_2^{'}(x^{(n)})}. \tag{A.20}$$

According to local convergence theory in [67, Theory 1.1], the standard assumptions for the convergence of (A.20) are [67, Assumption 1.2.1]

1. $F_2(x) = P_I l_{21}$ has a solution $x^\star$.
2. $F_2^{'}(x)$ is Lipschitz continuous near $x^\star$.
3. $F_2^{'}(x^\star)$ is nonsingular.

In the following, we prove that the sequence generated by (A.20) satisfies the standard assumptions.

Before proceeding to the proof, we first provide some useful expressions. The first, second, and the third derivative of $f_2(x)$ in (4.64) w.r.t. x are given as

$$f_2^{'}(x) = e^{\frac{-\lambda_b}{x}} \tag{A.21}$$

$$f_2^{''}(x) = e^{\frac{-\lambda_b}{x}} \frac{\lambda_b}{x^2} \tag{A.22}$$

$$f_2^{'''}(x) = e^{\frac{-\lambda_b}{x}} x^{-4} \lambda_b(\lambda_b - 2x) \tag{A.23}$$

and $f_2^{'}(x) = f_2^{''}(x) = f_2^{'''}(x) = 0$, $x \leq 0$.

Considering the first condition, we first check

$$F_2(0) - P_I l_{21} = -P_I l_{21} < 0 \tag{A.24}$$

and we have

$$\begin{aligned} \lim_{x\to\infty} F_2(x) - P_I l_{21} &= \lim_{x\to\infty} (f_2(x) - f_2(x - P_S)) - P_I l_{21} \\ &\overset{(a)}{>} \lim_{x\to\infty} P_S f_2^{'}(x - P_S) - P_I l_{21} \overset{(b)}{=} P_S - P_I l_{21} \overset{(c)}{>} 0 \end{aligned} \tag{A.25}$$

where, in (a), we use the property that $f_2(x)$ is convex due to the positivity of $f_2^{''}(x)$ in (A.21). In (b), we use $\lim_{x\to\infty} f^{'}(x) = 1$. The last step (c) uses the property that in

this case the interference power constraint is stricter than the peak power constraint, i.e., $P_S > P_I l_{21}$. Furthermore, we check the first derivative of $F_2(x)$:

$$F_2^{'}(x) = f_2^{'}(x) - f_2^{'}(x - P_S) \overset{(a)}{>} 0 \tag{A.26}$$

where in (a), we use the property that $f_2^{'}(x)$ is a monotonically increasing function due to the positivity of $f_2^{''}(x)$ in (A.21). From (A.26) it can be seen that $F_2(x)$ is a strictly monotonically increasing function of x. Combining the conditions that $F_2(0) - P_I l_{21}$ and $\lim_{x\to\infty} F_2(x) - P_I l_{21}$ have opposite signs, we conclude that the first condition holds and the solution is unique.

Considering the second condition, we recall that an everywhere differentiable function is Lipschitz continuous if and only if its first derivative is bounded. Therefore, the second condition holds if we prove $|F_2^{''}(x)| = \left|f_2^{''}(x) - f_2^{''}(x - P_S)\right|$ is bounded. The proof is as follows. According to $f_2^{'''}(x)$ in (A.21), the function $f_2^{''}(x)$ is monotonically increasing when $0 < x < \lambda_b/2$ and monotonically decreasing afterwards. Therefore, $f_2^{''}(x)$ is a unimodal function and has the maximum at $x = \lambda_b/2$. Consequently,

$$|F_2^{''}(x)| < 2f_2^{''}(x)\Big|_{x=\lambda_b/2} = \frac{8}{\lambda_b e^2} < \infty. \tag{A.27}$$

The third condition follows directly from $F_2^{'}(x) \neq 0$ according to (A.26).

So far, we have completed the proof that (A.20) satisfies the standard assumptions. The convergence thus depends on the selection of the initial point $x^{(0)}$. Herein, the approximation of the root of $F_2(x) = P_I l_{21}$ is provided as $x^{(0)}$. Specifically, using the convexity of $f(x)$, we have $F_2(x) = f_2(x) - f_2(x - P_S) < P_S f_2^{'}(x) = P_S e^{-\frac{\lambda_b}{x}}$. Let $P_S e^{-\frac{\lambda_b}{x^{(0)}}} = P_I l_{21}$, we obtain $x^{(0)}$ in the form of (4.66) with $\lambda_b = (a_1 + 1)l_{22}/b_1$.

A.7 Proof of Proposition 4.2.4

Replacing $P_2(g_{22})$ in (4.32) with $\tilde{P}_2^{\star}(g_{22})$ in (4.60) and (4.61), we have an expression of the achievable performance of the suboptimal strategy DT-WF. We treat the cases that $P_2(g_{22})$ is independent or dependent of g_{22} separately by defining the following two functions

$$\tilde{r}_1(\lambda, p) = \int_{\lambda}^{\infty} \left(\ln\left(1 + \frac{P_2(x)x}{\sigma_s^2}\right) + e^{l_{12}s_1} E_1(l_{12}s_1) - e^{l_{12}s_0} E_1(l_{12}s_0)\right) f_{g_{22}}(x)dx\Big|_{P_2(x)=p} \tag{A.28}$$

$$\tilde{r}_2(\lambda) = \int_{\lambda}^{\infty} \left(\ln\left(1 + \frac{P_2(x)x}{\sigma_s^2}\right) + e^{l_{12}s_1} E_1(l_{12}s_1) - e^{l_{12}s_0} E_1(l_{12}s_0)\right) f_{g_{22}}(x)dx\Big|_{P_2(x)=\tilde{p}^{\star}(x)} \tag{A.29}$$

with nonnegative λ and p. Note that s_1 is a function of $P_2(x)$. After some mathematical manipulations, we obtain

$$\tilde{R}_{\text{DT-WF}} = \begin{cases} \tilde{r}_1\left(g_1,\, P_S\right) - \tilde{r}_2\left(g_1\right) + \tilde{r}_2\left(g_0\right), & v_2^\star > P_S \\ \tilde{r}_2\left(g_0\right), & 0 < v_2^\star \leq P_S. \end{cases} \tag{A.30}$$

In order to get $\tilde{R}_{\text{DT-WF}}$ in closed form, we derive $\tilde{r}_1(\lambda, p)$ and $\tilde{r}_2(\lambda)$ in closed form. First, we decompose $\tilde{r}_1(\lambda, p)$ into three parts and evaluate them individually

$$\tilde{r}_1(\lambda, p) = \tilde{r}_{11}(\lambda, p) + \tilde{r}_{12}(\lambda, p) - \tilde{r}_{13}(\lambda) \tag{A.31}$$

where $\tilde{r}_{11}(\lambda,\, p)$ is given as

$$\begin{aligned} \tilde{r}_{11}(\lambda,\, p) &= \int_\lambda^\infty \ln\left(1 + \frac{px}{\sigma_s^2}\right) l_{22} e^{-l_{22}x} dx \\ &\overset{(a)}{=} \ln\left(1 + \frac{p\lambda}{\sigma_s^2}\right) e^{-l_{22}\lambda} + e^{\frac{l_{22}\sigma_s^2}{p}} E_1\left(l_{22}\left(\lambda + \frac{\sigma_s^2}{p}\right)\right). \end{aligned} \tag{A.32}$$

and in (a), we use integration by parts.

The second term $\tilde{r}_{12}(\lambda,\, p)$ is represented as

$$\begin{aligned} \tilde{r}_{12}(\lambda,\, p) &= \int_\lambda^\infty e^{l_{12}s_1} E_1(l_{12}s_1) l_{22} e^{-l_{22}x} dx \\ &\overset{(a)}{=} \int_\lambda^\infty e^{l_{12}\left(\frac{\sigma_s^2 + px}{P_1}\right)} E_1\left(l_{12}\left(\frac{\sigma_s^2 + px}{P_1}\right)\right) l_{22} e^{-l_{22}x} dx \\ &\overset{(b)}{=} \int_\lambda^\infty e^{a_2 + b_2 x} E_1\left(a_2 + b_2 x\right) l_{22} e^{-l_{22}x} dx \\ &\overset{(c)}{=} \frac{l_{22}}{b_2} e^{\frac{l_{22}a_2}{b_2}} \int_{a_2 + b_2\lambda}^\infty e^{\left(1 - \frac{l_{22}}{b_2}\right)t} E_1(t) dt \\ &\overset{(d)}{=} \frac{l_{22}}{b_2} e^{\frac{l_{22}a_2}{b_2}} f_3\left(a_2 + b_2\lambda, 1 - \frac{l_{22}}{b_2}\right) \end{aligned} \tag{A.33}$$

where, in (a), we take $P_2(x) = p$ into s_1 defined in (4.31), respectively. In (b), we use

$$a_2 = \frac{l_{12}\sigma_s^2}{P_1} \tag{A.34}$$

$$b_2 = \frac{l_{12}p}{P_1}. \tag{A.35}$$

We use $t = a_2 + b_2 x$ in (c). In (d), $f_3(\theta, \rho)$ is exploited which is defined and calculated using integration by parts as follows:

$$f_3(\theta, \rho) = \int_\theta^\infty e^{\rho x} E_1(x) dx, \qquad \theta \geq 0,\ \rho < 1 \tag{A.36}$$

$$= \begin{cases} \frac{1}{\rho}\left(-e^{\theta\rho}E_1(\theta) + E_1((1-\rho)\theta)\right), & \rho < 1 \text{ and } \rho \neq 0 \\ e^{-\theta} - \theta E_1(\theta), & \rho = 0. \end{cases} \tag{A.37}$$

Using the property that $s_0 = \sigma_s^2/P_1$ is independent of g_{22} according to (4.31), the third term is

$$\tilde{r}_{13}(\lambda) = \int_{\lambda}^{\infty} e^{l_{12}s_0} E_1(l_{12}s_0) l_{22} e^{-l_{22}x} dx = e^{\frac{l_{12}\sigma_s^2}{P_1} - l_{22}\lambda} E_1\left(\frac{l_{12}\sigma_s^2}{P_1}\right) \tag{A.38}$$

Second, we similarly decompose $\tilde{r}_2(\lambda)$ into three parts and evaluate them separately:

$$\tilde{r}_2(\lambda) = \tilde{r}_{21}(\lambda) + \tilde{r}_{22}(\lambda) - \tilde{r}_{23}(\lambda). \tag{A.39}$$

Unlike $P_2(g_{22})$ in $\tilde{r}_1(\lambda, p)$, $P_2(g_{22})$ is a function of g_{22} in $\tilde{r}_2(\lambda)$.

Particularly, the first term $\tilde{r}_{21}(\lambda)$ can be reformulated as

$$\begin{aligned} \tilde{r}_{21}(\lambda) &\overset{(a)}{=} \int_{\lambda}^{\infty} \ln\left(1 + \frac{\left(v_2^\star - \frac{a_1+1}{b_1 x}\right)x}{\sigma_s^2}\right) l_{22} e^{-l_{22}x} dx \\ &= \int_{\lambda}^{\infty} \ln\left(-\frac{P_1}{\sigma_s^2 l_{12}} + \frac{v_2^\star}{\sigma_s^2} x\right) l_{22} e^{-l_{22}x} dx \\ &\overset{(b)}{=} \ln\left(-\frac{P_1}{\sigma_s^2 l_{12}} + \frac{v_2^\star \lambda}{\sigma_s^2}\right) e^{-l_{22}\lambda} + e^{-\frac{P_1 l_{22}}{v_2^\star l_{12}}} E_1\left(l_{22}\left(\lambda - \frac{P_1}{v_2^\star l_{12}}\right)\right) \end{aligned} \tag{A.40}$$

where, in (a), we use the definition of $\tilde{p}_2^\star(g_{22})$ in (4.59). In (b), we use integration by parts. According to the definition of the exponential integral, the real part of the argument of $E_1(\cdot)$ should be positive. Therefore, we require that $\lambda > P_1/(v_2^\star l_{12})$ in (A.40). This is validated by noting that in this paper λ is chosen to be either g_1 or g_0 as shown in (4.69) and it is easy to verify that $g_1 > g_0 > P_1/(v_2^\star l_{12})$.

The second term $\tilde{r}_{22}$ is stated as

$$\begin{aligned} \tilde{r}_{22}(\lambda) &= \int_{\lambda}^{\infty} e^{l_{12}s_1} E_1(l_{12}s_1) l_{22} e^{-l_{22}x} dx \\ &\overset{(a)}{=} \int_{\lambda}^{\infty} e^{l_{12}\left(-\frac{1}{l_{12}} + \frac{v_2^\star}{P_1}x\right)} E_1\left(l_{12}\left(-\frac{1}{l_{12}} + \frac{v_2^\star}{P_1}x\right)\right) l_{22} e^{-l_{22}x} dx \\ &\overset{(b)}{=} \int_{\lambda}^{\infty} e^{a_3 + b_3 x} E_1\left(a_3 + b_3 x\right) l_{22} e^{-l_{22}x} dx \\ &\overset{(c)}{=} \frac{l_{22}}{b_2} e^{\frac{l_{22}a_3}{b_3}} f_3\left(a_3 + b_3\lambda, 1 - \frac{l_{22}}{b_3}\right) \end{aligned} \tag{A.41}$$

where, in (a), we take the definition of $\tilde{p}^\star(g_{22})$ in (4.59) into s_1 in (4.31). In (b), we use $a_3 = -1$ and $b_3 = v_2^\star l_{12}/P_1$. Step (c) is similar to (c) and (d) in (A.33).

Lastly, due to the independence of s_0 and g_{22}, it straightforwardly follows that

$$\tilde{r}_{23}(\lambda) = \tilde{r}_{13}(\lambda) = e^{\frac{l_{12}\sigma_s^2}{P_1} - l_{22}\lambda} E_1\left(\frac{l_{12}\sigma_s^2}{P_1}\right) \tag{A.42}$$

To summarize, in (4.69), $\tilde{r}_1(\lambda,\, p)$ is given by (A.31) in which $\tilde{r}_{11}(\lambda,\, p)$, $\tilde{r}_{12}(\lambda,\, p)$, and $\tilde{r}_{13}(\lambda,\, p)$ are provided in (A.32), (A.33), and (A.38), respectively. Similarly, $\tilde{r}_2(\lambda,\, p)$ is expressed in (A.39) in which $\tilde{r}_{21}(\lambda,\, p)$, $\tilde{r}_{22}(\lambda,\, p)$, and $\tilde{r}_{23}(\lambda,\, p)$ are shown in (A.40), (A.41), and (A.42), respectively.

A.8 Strong Duality for the Optimization Problem (4.93)

According to [138, Theorem 1], [127, Theorem 4.1], if an optimization problem satisfies both the time-sharing property [138, Definition 1] and the Slater regularity condition [111], the primal and the dual problem have zero duality gap. Thus, in order to prove strong duality for (4.93), we investigate these two conditions separately.

1) *Time-sharing property*: We assume $\check{P}_2^\star(t)$ and $\hat{P}_2^\star(t)$ are optimal solutions to (4.93) with ε equal to $\check{\varepsilon}$ and $\hat{\varepsilon}$, respectively. Similar to [138], we choose the solution $\tilde{P}_2(t)$ which is equal to $\check{P}_2^\star(t)$ for α of the channel realizations and equal to $\hat{P}_2^\star(t)$ for $1-\alpha$ of the channel realizations. It is straightforward to show that $\tilde{P}_2(t)$ is a feasible solution to (4.93) with

$$\varepsilon_p = \alpha\check{\varepsilon}_p + (1-\alpha)\hat{\varepsilon}_p$$

and that the following inequality

$$R\left(\tilde{P}_2(t)\right) \geq \alpha R\left(\check{P}_2^\star(t)\right) + (1-\alpha) R\left(\hat{P}_2^\star(t)\right) \tag{A.43}$$

is satisfied with equality. The term $R(\cdot)$ is defined in (4.9).

2) *Slater regularity condition*: For (4.93), the Slater regularity condition requires the existence of a solution in the relative interior of the domain that is strictly feasible [111]. It is easy to show that this condition holds for any $\varepsilon \in [0,1)$ since the solution $P_2^\star(t) = \delta$, $\delta \in \left(0, \frac{1}{a}(\frac{1}{\varepsilon} - 1)\right)$, $\forall t$, always satisfies all constraints with strict inequality.

A.9 Proof of Proposition 4.3.1

The Lagrangian of (4.93) is written as

$$L = \mathbb{E}_t\left\{\ln\left(1 + \frac{P_2(t)}{t}\right)\right\} + \lambda\left(\mathbb{E}_t\left\{\frac{1}{1+aP_2(t)}\right\} - \varepsilon\right) \tag{A.44}$$

where the non-negative Lagrangian multiplier λ corresponds to the constraint (4.91). Herein, no Lagrange multiplier is assigned to the peak power constraints since they are considered in the generalized KKT conditions as shown later.

According to the generalized KKT conditions for functional optimization [85,87], if the optimal solution $P_2^\star(t)$ is a regular point, it satisfies the following conditions

$$\left.\frac{dl}{dP_2(t)}\right|_{P_2(t)=P_2^\star(t)} \begin{cases} = 0, & 0 < P_2^\star(t) < P_S \\ \leq 0, & P_2^\star(t) = 0 \\ \geq 0, & P_2^\star(t) = P_S \end{cases} \tag{A.45}$$

$$\lambda^\star \geq 0 \tag{A.46}$$

$$\lambda^\star \left(\mathbb{E}_t \left\{ \frac{1}{1+aP_2^\star(t)} \right\} - \varepsilon \right) = 0 \tag{A.47}$$

$$\mathbb{E}_t \left\{ \frac{1}{1+aP_2^\star(t)} \right\} \geq \varepsilon \tag{A.48}$$

where the function l is defined as

$$l = \ln\left(1 + \frac{P_2(t)}{t}\right) + \lambda^\star \left(\frac{1}{1+aP_2(t)} - \varepsilon \right) \tag{A.49}$$

and the derivative of (A.45) with respect to (w.r.t.) $P_2(t)$ is

$$\frac{dl}{dP_2(t)} = \frac{1}{t+P_2(t)} - \lambda^\star \frac{a}{(1+aP_2(t))^2}. \tag{A.50}$$

By solving the KKT conditions (A.45), (A.46), and using $v = 1/\lambda$, we obtain the solutions as given in (4.95), (4.96), and (4.97).

The KKT conditions (A.47) and (A.48) are used to determine the value of $v^\star$. Particularly, if the outage constraint is inactive at the optimal solution, $v^\star$ should be infinity since $\lambda^\star = 0$ according to (A.47). We prove by contradiction that under this circumstance, the optimal solution is $P_2^\star(t) = P_S$ for all t. If $P_2^\star(t) \neq P_S$ for some t, the instantaneous transmit power could be increased to improve the secondary transmission rate while still satisfying the outage constraint. It can be seen from (A.48) that this case happens if $\varepsilon \in [0, 1/(1+aP_S))$. Furthermore, if $\varepsilon = 1/(1+aP_S)$, it is easy to see that the optimal solution is also $P_2^\star(t) = P_S$, thus the calculation of $v^\star$ is not required. Therefore, we only focus on calculating $v^\star$ if $1/(1+aP_S) < \varepsilon < 1$. In this case, $v^\star$ should be chosen to let the constraint $Q_c(P_2^\star(t)) = \varepsilon$ hold.

A.10 Feasible Region $\mathcal{F}_t$ for $x_{1,2}(t,v)$

The feasible region $\mathcal{F}_t$ is the set of t to let the solutions $x_{1,2}(t,v)$ be a real number with $0 < x_{1,2}(t,v) < P_S$. In order to guarantee that $x_{1,2}(t,v)$ is real, the non-negativity of the part under the square root in (4.98) and (4.99) is required. This yields

$$t \geq T_3(v), \quad T_3(v) = \frac{4v-1}{4av}. \tag{A.51}$$

In the following, we characterize the feasible regions of t for $x_1(t,v)$ and $x_2(t,v)$ separately. Firstly, considering $x_1(t,v)$, we distinguish between three cases:

1. $1-2v \leq 0$: Calculating the region of t to let $x_1(t,v)$ satisfy $0 < x_1(t,v) < P_S$, we have
$$T_2(v) < t < T_1(v) \tag{A.52}$$
where
$$T_1(v) = \frac{(1+aP_S)^2}{a}v - P_S, \quad T_2(v) = \frac{v}{a}, \tag{A.53}$$
and $T_1(v) > T_2(v)$. Additionally, we have $T_2(v) \geq T_3(v)$ since
$$T_2(v) - T_3(v) = \frac{1+4v^2-4v}{4av} = \frac{(1-2v)^2}{4av} \geq 0 \tag{A.54}$$
which shows that (A.52) implies the condition (A.51).

2. $0 < 1-2v < 2avP_S$: Given that the condition in (A.51) holds, the positivity of $x_1(t,v)$ is straightforwardly satisfied. We calculate the region of t to let $x_1(t,v) \in (0, P_S)$ hold and obtain $t < T_1(v)$. Therefore, the feasible region of t is
$$T_3(v) \leq t < T_1(v) \tag{A.55}$$
where we use the relation $T_1(v) \geq T_3(v)$ due to
$$T_1(v) - T_3(v) = \frac{(2v(aP_S+1)-1)^2}{4av} \geq 0. \tag{A.56}$$

3. $1-2v \geq 2avP_S$: It is easy to verify that $x_1(t,v)$ is not the feasible solution for any non-negative t since it will always be equal to or exceed P_S.

Secondly, we consider the solution $x_2(t,v)$. The term $1-2v$ should be positive in order to satisfy the positivity of $x_2(t,v)$. We classify the discussion into two cases:

1) $0 < 1-2v < 2avP_S$: If the condition in (A.51) holds, $x_2(t,v)$ is real and $x_2(t,v) < P_S$. To restrain $x_2(t,v)$ to be positive, we have $t < T_2(v)$. Therefore, the feasible region of t is
$$T_3(v) \leq t < T_2(v). \tag{A.57}$$

2) $1-2v \geq 2avP_S$: We obtain the region of t such that $0 < x_2(t,v) < P_S$ is satisfied as
$$T_1(v) < t < T_2(v) \tag{A.58}$$
where we use $T_1(v) < T_2(v)$ for $1-2v > avP_S$. Using the inequality $T_1(v) \geq T_3(v)$, cf. (A.56), we show that (A.58) implies the satisfaction of the condition (A.51).

In summary, the region of t to make $x_1(t,v)$ feasible is (A.52) for $1-2v \leq 0$, and (A.55) for $0 < 1-2v < 2avP_S$. The region of t to make $x_2(t,v)$ feasible is (A.57) for $0 < 1-2v < 2avP_S$ and (A.58) for $1-2v \geq 2avP_S$.

A.11 Proof of Lemma 4.3.1

Consider that for the case $1/(1+aP_S) < \varepsilon < 1$, the outage probability constraint holds with equality at the optimum and $v^\star$ is a positive finite value. Since the considered functions are in twice differentiable class, according to [58, Chapter 7.2, Proposition 1], a sufficient condition for $x^\star(t)$ to yield a local maximum of (4.93) is that $x^\star(t)$ is a regular point and satisfies

$$L_{xx}\left(x^\star(t), \lambda^\star; \xi(t)\right) < 0. \tag{A.59}$$

where $L_{xx}(\cdot;\cdot)$ is the second-order functional derivative of the Lagrange function L in (A.44) w.r.t. x in the direction $\xi(t)$. Similarly, a sufficient condition for $x^\star(t)$ yielding a local minimum of (4.93) is $L_{xx}\left(x^\star(t), \lambda^\star; \xi(t)\right) > 0$.

Without loss of generality, we focus on checking the sufficient conditions for a local maximum unless specified otherwise. Specifically, defining $Q_{\mathrm{c}}^{'}\left(x(t)\right)$ as the functional derivative of $Q_{\mathrm{c}}(x(t))$ w.r.t. $x(t)$ at $x^\star(t)$ [34], $\xi(t)$ is any vector of the kernel of the mapping $Q_{\mathrm{c}}^{'}(x^\star(t))$. Incorporating (A.44) into the left side of (A.59), we have

$$\begin{aligned} L_{xx}\left(x^\star(t), \lambda^\star; \xi(t)\right) &= \frac{d^2}{d\varphi^2} L\left(x^\star(t) + \varphi\xi(t), \lambda^\star\right)\Big|_{\varphi=0} \\ &= \int_0^\infty \left(-\frac{1}{\left(t + x^\star(t)\right)^2} + \frac{2\lambda^\star a^2}{\left(1 + a x^\star(t)\right)^3} \right) f(t)\xi^2(t)dt \end{aligned} \tag{A.60}$$

with $f(t)$ representing the PDF of t. Integrating (A.60) into (A.59), it yields

$$G(x^\star(t)) = -\frac{1}{\left(t + x^\star(t)\right)^2} + \frac{2\lambda^\star a^2}{\left(1 + a x^\star(t)\right)^3} < 0. \tag{A.61}$$

However, verifying the optimality of $x_1(t, v^\star)$ and $x_2(t, v^\star)$ by checking (A.61) is difficult since incorporating them into (A.61) yields involved expressions. Alternatively, we rewrite (A.61) into equivalent form that is simple to verify. Specifically, using $\lambda = 1/v$, we construct

$$\bar{x}_{1,2}(t, \lambda) = \frac{\lambda - 2 \pm \sqrt{\lambda^2 - 4\lambda + 4a\lambda t}}{2a} \tag{A.62}$$

and it is easy to verify that

$$\bar{x}_1(t, \lambda) = x_1(t, v),\ \bar{x}_2(t, \lambda) = x_2(t, v).$$

Consequently, $\bar{x}_1(t, \lambda^\star)$ and $\bar{x}_2(t, \lambda^\star)$ are roots of $dl/dP_2(t) = 0$, where $dl/dP_2(t)$ was given in (A.50). Thus, the equation

$$\begin{aligned} H(x)\Big|_{x=\bar{x}_{1,2}(t,\lambda^\star)} &= \frac{\left(1 + ax\right)^2}{t + x}\Big|_{x=\bar{x}_{1,2}(t,\lambda^\star)} \\ &= \lambda^\star a \end{aligned} \tag{A.63}$$

follows directly. With the help of (A.63), $G(x^\star(t))$ in (A.61) at $\bar{x}_{1,2}(t,\lambda^\star)$ is

$$\begin{aligned} G(\bar{x}_{1,2}(t,\lambda^\star)) &\overset{(a)}{=} -\frac{1}{(t+\bar{x}_{1,2}(t,\lambda^\star))^2}+\frac{1}{t+\bar{x}_{1,2}(t,\lambda^\star)}\frac{2a}{(1+a\bar{x}_{1,2}(t,\lambda^\star))} \\ &= \frac{1}{(1+a\bar{x}_{1,2}(t,\lambda^\star))^2}\left(-\frac{(1+a\bar{x}_{1,2}(t,\lambda^\star))^2}{(t+\bar{x}_{1,2}(t,\lambda^\star))^2}+\frac{2a(1+a\bar{x}_{1,2}(t,\lambda^\star))}{t+\bar{x}_{1,2}(t,\lambda^\star)}\right) \\ &\overset{(b)}{=} \frac{1}{(1+a\bar{x}_{1,2}(t,\lambda^\star))^2}H'(\bar{x}_{1,2}(t,\lambda^\star)) \end{aligned} \tag{A.64}$$

where, in (a), we incorporate (A.63) and, in (b), we insert $H^{'}(x)$ at $\bar{x}_{1,2}(t,\lambda^\star)$. Comparing (A.64) with (A.61), we conclude that in order to satisfy the sufficient condition (A.61) for a local maximum point, we can equivalently check whether $H'(x)|_{x=\bar{x}_{1,2}(t,\lambda^\star)} < 0$. This condition combined with $H(\bar{x}_{1,2}(t,\lambda^\star)) = \lambda^\star a$ implies that if $\bar{x}_{1,2}(t,\lambda^\star)$ is a strictly monotonically decreasing function of $\lambda^\star$, then $\bar{x}_{1,2}(t,\lambda^\star)$ is a local maximum point. Analogously, if $\bar{x}_{1,2}(t,\lambda^\star)$ is a strictly monotonically increasing function of $\lambda^\star$, then $\bar{x}_{1,2}(t,\lambda^\star)$ is a local minimum point.

This equivalent sufficient condition is verified as follows. Taking the derivative of $\bar{x}_{1,2}(t,\lambda^\star)$ in (4.98) and (4.99) w.r.t. to $\lambda^\star$ yields

$$\frac{d\bar{x}_{1,2}(t,\lambda^\star)}{d\lambda^\star} = \frac{1}{2a}\left(1 \pm \frac{\lambda^\star - 2 + 2at}{\sqrt{(\lambda^\star)^2 - 4\lambda^\star + 4at\lambda^\star}}\right) \tag{A.65}$$

We use the property that the term under the square root should be non-negative, i.e., $(\lambda^\star)^2 - 4\lambda^\star + 4at\lambda^\star \geq 0$, we thus have $t \geq \left(4\lambda^\star - (\lambda^\star)^2\right)/(4a\lambda^\star)$. Consequently, we have

$$\lambda^\star - 2 + 2at \geq \lambda^\star - 2 + \frac{4\lambda^\star - (\lambda^\star)^2}{2\lambda^\star} = \frac{\lambda^\star}{2} \overset{(a)}{>} 0 \tag{A.66}$$

where, in (a), we use that herein $\lambda^\star$ is positive. Taking it into (A.65), for $x_1^\star(t)$

$$\frac{d\bar{x}_1(t,\lambda^\star)}{d\lambda^\star} = \frac{1}{2a}\left(1 + \frac{\lambda^\star - 2 + 2at}{\sqrt{(\lambda^\star)^2 - 4\lambda^\star + 4at\lambda^\star}}\right) > 0 \tag{A.67}$$

For $x_2^\star(t)$, we obtain

$$\frac{d\bar{x}_2(t,\lambda^\star)}{d\lambda^\star} \overset{(a)}{<} \frac{1}{2a}\left(1 - \frac{\lambda^\star - 2 + 2at}{\lambda^\star - 2 + 2at}\right) = 0 \tag{A.68}$$

where, in (a), we use the positivity of $\lambda^\star - 2 + 2at$ in (A.66) and the inequality

$$(\lambda^\star)^2 - 4\lambda^\star + 4at\lambda^\star < (\lambda^\star - 2 + 2at)^2 \tag{A.69}$$

for $at \neq 1$. Note that we do not consider the case when $at = 1$ since incorporating it into (A.62) yielding $\bar{x}_2(t,\lambda^\star) = -1/a$, which is negative and infeasible.

The inequalities (A.67) and (A.68) imply that $\bar{x}_1(t,\lambda^\star)$ and $\bar{x}_2(t,\lambda^\star))$ are strictly monotonically increasing and decreasing functions of $\lambda^\star$, respectively. Therefore,

$x_1(t,v)$ is a local minimum solution and $x_2(t,v)$ is a local maximum solution in the corresponding region of t.

A.12 Proof of Lemma 4.3.2

In this Appendix, we use $x^\star(t)$ to denote the optimal solution of (4.93). The lemma is proved by reduction to absurdity. Specifically, assuming there exists a t_1 and a t_2 with $t_1 < t_2$ such that $x^\star(t)$ in the vicinity of t_1 and t_2 is continuous and values of $x^\star(t)$ in the vicinity of t_1 are strictly smaller than those in the vicinity of t_2. In the remaining region of t, $x^\star(t)$ is assumed to be decreasing w.r.t. t. We choose small δ_1 and δ_2 satisfying $t_1+\delta_1 < t_2-\delta_2$ and

$$A=\int_{t_1-\delta_1}^{t_1+\delta_1} f(t)dt=\int_{t_2-\delta_2}^{t_2+\delta_2} f(t)dt.$$

According to the first mean value theorem [52] and using the property that the pdf $f(t)$ is bounded, there exist $\tilde{t}_1$ and $\tilde{t}_2$ such that

$$\begin{aligned}\int_{t_1-\delta_1}^{t_1+\delta_1}\frac{1}{1+ax^\star(t)}f(t)dt&=\frac{1}{1+ax^\star(\tilde{t}_1)}A\\ \int_{t_2-\delta_2}^{t_2+\delta_2}\frac{1}{1+ax^\star(t)}f(t)dt&=\frac{1}{1+ax^\star(\tilde{t}_2)}A.\end{aligned}$$

Note that $x^\star(\tilde{t}_1) < x^\star(\tilde{t}_2)$. We can construct a new solution

$$\bar{x}^\star(t) \;=\begin{cases}x^\star(\tilde{t}_2), & \text{if } t\in[t_1-\delta_1,\ t_1+\delta_1]\\ x^\star(\tilde{t}_1), & \text{if } t\in[t_2-\delta_2,\ t_2+\delta_2]\\ x^\star(t), & \text{otherwise}\end{cases}. \tag{A.70}$$

The solution $\bar{x}^\star(t)$ yields the same outage probability for the PU as $x^\star(t)$. Therefore, it also satisfies the outage probability constraint. Calculating the difference between the secondary achievable rate with $\bar{x}^\star(t)$ and that with $x^\star(t)$, we obtain

$$\begin{aligned}&\lim_{\delta_1,\delta_2\to0} R(\bar{x}^\star(t))-R(x^\star(t))\\ =&\lim_{\delta_1\to0}\int_{t_1-\delta_1}^{t_1+\delta_1}\left(\ln\left(1+\frac{x^\star(\tilde{t}_2)}{t}\right)-\ln\left(1+\frac{x^\star(t)}{t}\right)\right)f(t)dt\\ &+\lim_{\delta_2\to0}\int_{t_2-\delta_2}^{t_2+\delta_2}\left(\ln\left(1+\frac{x^\star(\tilde{t}_1)}{t}\right)-\ln\left(1+\frac{x^\star(t)}{t}\right)\right)f(t)dt\\ \overset{(a)}{=}&\lim_{\delta_1,\delta_2\to0}A\left(\ln\left(\frac{\bar{t}_1+x^\star(\tilde{t}_2)}{\bar{t}_1+x^\star(\bar{t}_1)}\right)+\ln\left(\frac{\bar{t}_2+x^\star(\tilde{t}_1)}{\bar{t}_2+x^\star(\bar{t}_2)}\right)\right)\\ \overset{(b)}{=}&A\left(\ln\left(\frac{\tilde{t}_1+x^\star(\tilde{t}_2)}{\tilde{t}_1+x^\star(\tilde{t}_1)}\right)+\ln\left(\frac{\tilde{t}_2+x^\star(\tilde{t}_1)}{\tilde{t}_2+x^\star(\tilde{t}_2)}\right)\right)\overset{(c)}{>}0\end{aligned} \tag{A.71}$$

where, in (a), the first mean value theorem is applied. In (b), we use the fact that $\lim_{\delta_1 \to 0} \bar{t}_1 = \tilde{t}_1$ and $\lim_{\delta_2 \to 0} \bar{t}_2 = \tilde{t}_2$. The last step (c) is due to the a priori assumption that $x^\star(\tilde{t}_1) < x^\star(\tilde{t}_2)$

The inequality (A.71) states that there exists another solution $\bar{x}^\star(t)$ that is a decreasing function of t that satisfies the outage constraint and results in a higher achievable rate than the solution $x^\star(t)$. Therefore, the optimal solution of (4.93) is a decreasing function of t.

A.13 Proof of Proposition 4.3.3

For the case $1 - 2\tilde{v} \geq 2a\tilde{v}P_S$, we incorporate (4.106) into $Q_c(\tilde{v})$ with $\tilde{x}_2(t, \tilde{v})$ in (4.111) and have

$$Q_c(\tilde{v}) = \frac{1}{1 + aP_S} \int_0^{T_1(\tilde{v})} f(t)dt + \frac{1 - 2\tilde{v}}{a} \int_{T_1(\tilde{v})}^{T_2(\tilde{v})} \frac{1}{(1-\tilde{v})/a - t} f(t)dt + \int_{T_2(\tilde{v})}^{\infty} f(t)dt$$

Defining

$$Q_1(T) = \int_0^T f(t)dt \tag{A.72}$$

$$Q_2(\lambda, T) = \int_0^T \frac{1}{\lambda - t} f(t)dt \tag{A.73}$$

with $\lambda > T$ and nonnegative λ, we arrive at (4.113). A similar derivation is performed to generate (4.114), (4.115) and (4.116). Therefore, the key to derive a closed-form expression of $Q_c(\tilde{v})$ is to obtain the analytical form of $Q_1(T(\tilde{v}))$ and $Q_2(T(\tilde{v}))$. Note that we use T instead of $T(\tilde{v})$ for notational brevity in the remainder of this proof.

It is easy to verify $Q_1(T) = 0$ and $Q_2(T) = 0$ for $T < 0$. In the following, we only consider the case $T \geq 0$. Taking the PDF of t (A.5) into (A.72) and (A.73), we have

$$\begin{aligned}
Q_1(T) &\overset{(a)}{=} b \int_\rho^\infty \frac{z + b + 1}{(z+b)^2} e^{-z} dz \\
&= b \left(\int_\rho^\infty \frac{1}{(z+b)} e^{-z} dz + \int_\rho^\infty \frac{1}{(z+b)^2} e^{-z} dz \right) \\
&\overset{(b)}{=} b \left(e^b E_1(b+\rho) + \frac{1}{b+\rho} e^{-\rho} - e^b E_1(b+\rho) \right) \\
&= \frac{b}{b+\rho} e^{-\rho}.
\end{aligned}$$

where, in (a), we use the substitutions to simplify the integral

$$b = \frac{l_{12}\sigma_S^2}{P_1},\ \rho = \frac{l_{22}\sigma_S^2}{T},\ z = \frac{l_{22}\sigma_S^2}{t} \tag{A.74}$$

In (b), we use the exponential integral $E_1(x) = \int_x^\infty e^{-t}/t dt$, $\mathcal{R}(x) > 0$ [4].

Similarly, we derive $Q_2(T)$ by taking (A.5) into its expression (A.73). After some mathematical computations, we have

$$\begin{aligned} Q_2(\lambda, T) &= \int_0^T \frac{1}{\lambda - t} f(t) dt \\ &= \frac{b}{\lambda} \int_d^\infty \frac{z}{z-c} \frac{z+b+1}{(z+b)^2} e^{-z} dz \\ &= \frac{b}{\lambda} \int_d^\infty \left(\left(\frac{c}{b+c} + \frac{c}{(b+c)^2} \right) \frac{1}{z-c} + \left(\frac{b}{b+c} - \frac{c}{(b+c)^2} \right) \frac{1}{z+b} \right. \\ &\quad \left. + \frac{b}{(b+c)(z+b)^2} \right) e^{-z} dz \\ &= \frac{b}{\lambda} \left(\frac{c(b+c+1)}{(b+c)^2} e^{-c} E_1(d-c) + \frac{b}{(b+c)(b+d)} e^{-d} \right. \\ &\quad \left. - \frac{c}{(b+c)^2} e^{b} E_1(b+d) \right) \end{aligned}$$

where $c = \frac{l_{22}\sigma_S^2}{\lambda}$.

A.14 Proof of Proposition 4.3.4

Applying the definition $r(T, \alpha, \beta)$ in (A.8) with $\alpha + \frac{\beta}{T} > 0$ to the achievable rate expressions, e.g., for the case $1 - 2\tilde{v}^\star \geq 2a\tilde{v}^\star P_S$, we incorporate (4.106) into (4.9) and have

$$R\left(\tilde{P}_2^\star(t)\right) = \int_0^{T_1(\tilde{v}^\star)} \ln\left(1 + \frac{P_S}{t}\right) f(t) dt + \int_{T_1(\tilde{v}^\star)}^{T_2(\tilde{v}^\star)} \ln\left(\frac{-2\tilde{v}^\star}{1 - 2\tilde{v}^\star} + \frac{\tilde{v}^\star}{(1-2\tilde{v}^\star)at} \right) f(t) dt \tag{A.75}$$

after some simple mathematical computations, we can rewrite (A.75) as (4.117). Similarly, by incorporating (4.107), (4.109), and (4.110) into (4.9), we represent the achievable rate as a function of $r(T, \alpha, \beta)$ in (4.118), (4.119), and (4.120), respectively.

A.15 Solution of $\Delta\widetilde{\mathbf{G}}_k$

We calculate the expression of $\Delta\widetilde{\mathbf{G}}_k$ in (5.65), for $\forall\, k = 1, \ldots, K$

$$\Delta\widetilde{\mathbf{G}}_k := \nabla_k^* \sum\nolimits_{l=1}^{K} \mathbb{E}_{\mathbf{H}_l} \{\mathrm{MSE}_l\} \tag{A.76}$$

$$
\begin{aligned}
&= \frac{\partial \sum_{l=1}^{K} \mathrm{Tr}\left\{\mathbf{I}_{N_l} - \widetilde{\mathbf{G}}_l^H \mathbf{L}\hat{\mathbf{H}}_l^H \bar{\mathbf{B}}_l^{-1} \hat{\mathbf{H}}_l \mathbf{L}^H \widetilde{\mathbf{G}}_l\right\}}{\partial \widetilde{\mathbf{G}}_k^*} \\
&= \underbrace{-\frac{\sum_{l=1}^{K} \mathrm{Tr}\left\{\partial(\widetilde{\mathbf{G}}_l^H) \mathbf{L}\hat{\mathbf{H}}_l^H \bar{\mathbf{B}}_l^{-1} \hat{\mathbf{H}}_l \mathbf{L}^H \widetilde{\mathbf{G}}_l\right\}}{\partial \widetilde{\mathbf{G}}_k^*}}_{L_1} \\
&\quad + \underbrace{\frac{\sum_{l=1}^{K} \mathrm{Tr}\left\{\widetilde{\mathbf{G}}_l^H \mathbf{L}\hat{\mathbf{H}}_l^H \bar{\mathbf{B}}_l^{-1} \partial\left(\bar{\mathbf{B}}_l\right) \bar{\mathbf{B}}_l^{-1} \hat{\mathbf{H}}_l \mathbf{L}^H \widetilde{\mathbf{G}}_l\right\}}{\partial \widetilde{\mathbf{G}}_k^*}}_{L_2}
\end{aligned}
$$

where L_1 and L_2 are the abbreviations of the two derivative terms. In the following we derive the detailed expressions of these terms. Specifically, for the first term

$$
\begin{aligned}
&\frac{\partial L_1}{\partial \left[\widetilde{\mathbf{G}}_k^*\right]_{m,n}} = \mathrm{Tr}\left\{\mathbf{e}_m \mathbf{e}_n^H \mathbf{L}\hat{\mathbf{H}}_k^H \bar{\mathbf{B}}_k \hat{\mathbf{H}}_k \mathbf{L}^H \widetilde{\mathbf{G}}_k\right\} = \mathrm{Tr}\left\{\mathbf{e}_n^H \mathbf{L}\hat{\mathbf{H}}_k^H \bar{\mathbf{B}}_k \hat{\mathbf{H}}_k \mathbf{L}^H \widetilde{\mathbf{G}}_k \mathbf{e}_m\right\} \\
&\Rightarrow \frac{\partial L_1}{\partial \widetilde{\mathbf{G}}_k^*} = \mathbf{L}\hat{\mathbf{H}}_k^H \bar{\mathbf{B}}_k \hat{\mathbf{H}}_k \mathbf{L}^H \widetilde{\mathbf{G}}_k
\end{aligned} \tag{A.77}
$$

For the second term, the derivative is given as

$$
\begin{aligned}
&\frac{\partial L_2}{\partial [\widetilde{\mathbf{G}}_k^*]_{m,n}} \\
&= \sum_{l=1}^{K} \mathrm{Tr}\Bigg\{ \widetilde{\mathbf{G}}_l^H \mathbf{L}\hat{\mathbf{H}}_l^H \bar{\mathbf{B}}_l^{-1} \frac{\partial \left(\sum_{i=1}^{K} \mathbf{H}_l \mathbf{L}^H \widetilde{\mathbf{G}}_i \widetilde{\mathbf{G}}_i^H \mathbf{L}\mathbf{H}_l^H + \mathrm{Tr}\left\{ \mathbf{L}^H \widetilde{\mathbf{G}}_i \widetilde{\mathbf{G}}_i^H \mathbf{L} \left(\mathbf{R}_{\mathbf{H}_i}^{e,Tx}\right)^T \right\} \mathbf{R}_{\mathbf{H}_i}^{e,Rx} \sigma_l^2 \mathbf{I} \right)}{\partial [\widetilde{\mathbf{G}}_k^*]_{mn}} \\
&\quad \bar{\mathbf{B}}_l^{-1} \hat{\mathbf{H}}_l \mathbf{L}^H \widetilde{\mathbf{G}}_l \Bigg\} \\
&= \sum_{l=1}^{K} \mathrm{Tr}\Bigg\{ \widetilde{\mathbf{G}}_l^H \mathbf{L}\hat{\mathbf{H}}_l^H \bar{\mathbf{B}}_l^{-1} \left(\hat{\mathbf{H}}_k \mathbf{L}^H \widetilde{\mathbf{G}}_k \mathbf{e}_m \mathbf{e}_n^H \mathbf{L}\hat{\mathbf{H}}_k^H + \mathrm{Tr}\left\{ \mathbf{L}^H \widetilde{\mathbf{G}}_k \mathbf{e}_m \mathbf{e}_n^H \mathbf{L} \left(\mathbf{R}_{\mathbf{H}_k}^{e,Tx}\right)^T \right\} \mathbf{R}_{\mathbf{H}_k}^{e,Rx} \right) \\
&\quad \bar{\mathbf{B}}_l^{-1} \hat{\mathbf{H}}_l \mathbf{L}^H \widetilde{\mathbf{G}}_l \Bigg\} \\
&= \sum_{l=1}^{K} \Bigg(\mathrm{Tr}\left\{ \mathbf{e}_n^H \mathbf{L}\hat{\mathbf{H}}_k^H \bar{\mathbf{B}}_l^{-1} \hat{\mathbf{H}}_k \mathbf{L}^H \widetilde{\mathbf{G}}_l \widetilde{\mathbf{G}}_l^H \mathbf{L}\hat{\mathbf{H}}_l^H \bar{\mathbf{B}}_l^{-1} \hat{\mathbf{H}}_k \mathbf{L}^H \widetilde{\mathbf{G}}_k \mathbf{e}_m \right\} \\
&\quad + \mathrm{Tr}\left\{ \mathbf{e}_n^H \mathbf{L} \left(\mathbf{R}_{\mathbf{H}_i}^{e,Tx}\right)^T \mathbf{L}^H \widetilde{\mathbf{G}}_k \mathbf{e}_m \right\} \mathrm{Tr}\left\{ \widetilde{\mathbf{G}}_l^H \mathbf{L}\hat{\mathbf{H}}_l^H \bar{\mathbf{B}}_l^{-1} \mathbf{R}_{\mathbf{H}_k}^{e,Rx} \bar{\mathbf{B}}_l^{-1} \hat{\mathbf{H}}_l \mathbf{L}^H \widetilde{\mathbf{G}}_l \right\} \Bigg).
\end{aligned} \tag{A.78}
$$

Thus, we have

$$\frac{\partial L_2}{\partial \widetilde{\mathbf{G}}_k^*} = \sum_{l=1}^{K} \Bigg(\mathbf{L}\hat{\mathbf{H}}_k^H \bar{\mathbf{B}}_l^{-1} \hat{\mathbf{H}}_l \mathbf{L}^H \widetilde{\mathbf{G}}_l \widetilde{\mathbf{G}}_l^H \mathbf{L} \hat{\mathbf{H}}_l^H \bar{\mathbf{B}}_l^{-1} \hat{\mathbf{H}}_k \mathbf{L}^H \widetilde{\mathbf{G}}_k + \mathrm{Tr}\left\{ \widetilde{\mathbf{G}}_l^H \mathbf{L} \hat{\mathbf{H}}_l^H \bar{\mathbf{B}}_l^{-1} \mathbf{R}_{\mathbf{H}_k}^{e,Rx} \bar{\mathbf{B}}_l^{-1} \hat{\mathbf{H}}_l \mathbf{L}^H \widetilde{\mathbf{G}}_l \right\} \mathbf{L} \left(\mathbf{R}_{\mathbf{H}_k}^{e,Tx} \right)^T \mathbf{L}^H \widetilde{\mathbf{G}}_k \Bigg). \tag{A.79}$$

Combining (A.77) and (A.79), we arrive at (5.65).

A.16 Solution of the Projection Operation

Consider the problem (5.66), both the objective function and the power constraint are convex functions of $\widetilde{\mathbf{G}}$. Thus, solving the projection operation is equivalent to solving a convex optimization problem. The Lagrange function of (5.66) is

$$L(\hat{\mathbf{G}}, \mu) = \sum_{k=1}^{K} \left\| \hat{\mathbf{G}}_k - \widetilde{\mathbf{G}}_k \right\|_F^2 + \mu \left(\sum_{k=1}^{K} \mathrm{Tr}\left\{ \hat{\mathbf{G}}_k \hat{\mathbf{G}}_k^H \right\} - P \right).$$

where μ is the Lagrange multiplier associated with the equivalent power constraint. According to KKT conditions, the following equality holds to derive the optimum

$$\begin{aligned} &\hat{\mathbf{G}}_k^\star - \widetilde{\mathbf{G}}_k + \mu \hat{\mathbf{G}}_k^\star = 0, \quad \forall k = 1, \ldots, K \\ &\sum_{k=1}^{K} \mathrm{Tr}\left\{ \hat{\mathbf{G}}_k^\star \left(\hat{\mathbf{G}}_k^\star \right)^H \right\} - P = 0 \end{aligned}$$

Solving from the last two equations, we obtain

$$\hat{\mathbf{G}}_k^\star = \sqrt{\frac{P}{\sum_{i=1}^{K} \left\| \widetilde{\mathbf{G}}_i \right\|_F^2}} \widetilde{\mathbf{G}}_k.$$

A.17 Solution of $g_{\mathbf{upp}}$

Defining the function

$$F(g_{22}) = \frac{p(\mathcal{H}_1) f(\gamma|\mathcal{H}_1)}{\frac{\sigma_s^2 + P_1/l_{12}}{g_{22}} + P_S} + \frac{p(\mathcal{H}_0) f(\gamma|\mathcal{H}_0)}{\frac{\sigma_s^2}{g_{22}} + P_S} \tag{A.80}$$

we note that it has the finite limit when g_{22} approaches infinity

$$\lim_{g_{22} \to \infty} F(g_{22}) = \frac{p(\mathcal{H}_0) f(\gamma|\mathcal{H}_0) + p(\mathcal{H}_1) f(\gamma|\mathcal{H}_1)}{P_S}. \tag{A.81}$$

Combined with (6.19), this indicates that if $\lim_{g_{22}\to\infty} F(g_{22}) \leq \lambda_2^\star f(\gamma|\mathcal{H}_1)$, then

$$g_{\text{upp}} = \infty \tag{A.82}$$

which means that the feasible condition (6.19) for $P_2^\star(\gamma, g_{22}) = P_S$ is always violated, i.e., $P_2^\star(\gamma, g_{22})$ can not reach the value P_S. Herein, we only consider the finite g_{22}.

If $\lim_{g_{22}\to\infty} F(g_{22}) > \lambda_2^\star f(\gamma|\mathcal{H}_1)$, we try to solve g_{upp} as the root of

$$F(x) = \lambda_2^\star f(\gamma|\mathcal{H}_1). \tag{A.83}$$

After some mathematical calculation, we reformulate (A.83) as a quadratic equation

$$a_1 x^2 + b_1 x + c_1 = 0 \tag{A.84}$$

with

$$a_1 = P_S \left(p(\mathcal{H}_0) f(\gamma|\mathcal{H}_0) + p(\mathcal{H}_1) f(\gamma|\mathcal{H}_1) - \lambda_2^\star f(\gamma|\mathcal{H}_1) P_S\right) \tag{A.85}$$

$$\begin{aligned} b_1 = {} & p(\mathcal{H}_1) f(\gamma|\mathcal{H}_1)\sigma_s^2 + p(\mathcal{H}_0) f(\gamma|\mathcal{H}_0)\left(\sigma_s^2 + \frac{P_1}{l_{12}}\right) \\ & - \lambda_2^\star f(\gamma|\mathcal{H}_1) P_S \left(2\sigma_s^2 + \frac{P_1}{l_{12}}\right) \end{aligned} \tag{A.86}$$

$$c_1 = -\lambda_2^\star f(\gamma|\mathcal{H}_1)\left(\sigma_s^4 + \frac{P_1}{l_{12}}\sigma_s^2\right). \tag{A.87}$$

Note that the preliminary condition $\lim_{g_{22}\to\infty} F(g_{22}) > \lambda_2^\star f(\gamma|\mathcal{H}_1)$ implies $a_1 > 0$. In addition, c_1 is negative and finite due to the positive and finite value of $\lambda_2^\star$.

The roots solving the quadratic equation (A.84) are

$$x_{1,2} = \frac{-b_1 \pm \sqrt{b_1^2 - 4a_1c_1}}{2a_1} \tag{A.88}$$

with $x_1 x_2 = c_1/a_1 < 0$. Using (A.85)-(A.87), we can verify that $\sqrt{b_1^2 - 4a_1c_1} > |b_1|$. Hence, we obtain g_{upp} as the nonnegative solution among $x_{1,2}$

$$g_{\text{upp}} = \frac{-b_1 + \sqrt{b_1^2 - 4a_1c_1}}{2a_1}. \tag{A.89}$$

To summarize, g_{upp} is given as

$$g_{\text{upp}} = \begin{cases} \dfrac{-b_1 + \sqrt{b_1^2 - 4a_1c_1}}{2a_1}, & \lim_{g_{22}\to\infty} F(g_{22}) > \lambda_2^\star f(\gamma|\mathcal{H}_1) \\ \infty, & \text{otherwise.} \end{cases} \tag{A.90}$$

Appendix B

Notation

$a, b, \ldots, A, B, \ldots$	scalars
$\mathbf{a}, \mathbf{b}, \ldots$	vectors
$\mathrm{diag}\{\mathbf{a}\}$	square matrix: the elements of $\mathbf{a}$ are on the diagonal and the other entries are all zero
$\mathbf{A}, \mathbf{B}, \ldots$	matrices
$\mathcal{A}, \mathcal{B}, \mathcal{C}, \ldots$	sets
$\lvert\mathcal{A}\rvert$	cardinality of the set $\mathcal{A}$
$\mathbb{N}$	set of all natural numbers
$\mathbb{C}$	set of all complex numbers
$[\mathbf{a}]_k$	kth element of the vector $\mathbf{a}$
$[\mathbf{A}]_{k,l}$	element of the matrix $\mathbf{A}$ in the kth row and lth column
$\mathbf{A}^T$	transpose of the matrix $\mathbf{A}$
$\mathbf{A}^H$	conjugate transpose of the matrix $\mathbf{A}$
$\mathbf{A}^*$	(element-wise) complex conjugate of the matrix $\mathbf{A}$
$\mathrm{vec}\{\mathbf{A}\}$	vectorization, i.e., column-wise stacking, of the matrix $\mathbf{A}$
$\mathrm{tr}\{\mathbf{A}\}$	trace of the square matrix $\mathbf{A}$
$\det\{\mathbf{A}\}$	determinant of the square matrix $\mathbf{A}$
$\lVert\mathbf{A}\rVert_F$	Frobenius norm of the matrix $\mathbf{A}$
$\mathrm{rank}\{\mathbf{A}\}$	rank of the matrix $\mathbf{A}$
$\lambda_{\max}(\mathbf{A})$	maximal eigenvalue of the Hermitian matrix $\mathbf{A}$
$\mathbf{A}^{\frac{1}{2}}$	unique Hermitian positive semidefinite square root of the Hermitian positive semidefinite matrix $\mathbf{A}$
$\mathbf{A} \odot \mathbf{B}$	Hadamard (element-wise) product of the matrices $\mathbf{A}$ and $\mathbf{B}$
$\mathbf{A} \otimes \mathbf{B}$	Kronecker product of the matrices $\mathbf{A}$ and $\mathbf{B}$
$\Pr\{\mathcal{A}\}$	probability of the event $\mathcal{A}$
$\mathbf{I}_N$	$N \times N$ identity matrix
$\mathbf{0}_{M,N}$	$M \times N$ all-zero matrix
$\mathbf{0}_N$	$N \times N$ all-zero matrix
$\mathbf{1}_{M,N}$	$M \times N$ all-one matrix
$\mathbf{e}_k$	kth standard basis for K-dimensional space
$x^{(n)}$	the value of x in the nth iteration
$\sum_k x[k]$	sum (from $-\infty$ to ∞ unless otherwise specified)
$\{x_k\}_{k=1}^K$	the set contains x_k, $k = 1, \ldots, K$
$a^\star$	optimal value of a
$\ln(a)$	natural logarithm of a
$\log_2(a)$	logarithm of a to the base 2

$\log_{10}(a)$	logarithm of a to the base 10
$(a)^+$	$\max(a, 0)$
$\mathbb{E}_x\{y(x)\}$	expectation of $y(x)$ over x
$f^{'}(x)$	first derivative of the function $f(x)$ over x
$f^{''}(x)$	second derivative of the function $f(x)$ over x
$f^{'''}(x)$	third derivative of the function $f(x)$ over x
$\mathcal{R}(x)$	real part of x
$\mathcal{CN}(\boldsymbol{\mu}, \mathbf{R})$	multivariate complex normal distribution of a k-dimensional random vector with location parameter $\boldsymbol{\mu}$ and covariance matrix $\mathbf{R}$

Acronyms

ALRT	approximated likelihood ratio test
AP	access point
AWGN	additive white Gaussian noise
BER	bit error rate
BPSK	binary phase-shift keying
BS	base station
CDF	cumulative density function
CR	cognitive radio
CSI	channel state information
DL-DA	downlink-based dual-loop algorithm
DoF	degree of freedom
DSA	dynamic spectrum access
DT-WF	double threshold waterfilling
DTCP-WF	double threshold constant-power waterfilling
ED	energy detection
ETSI	European Telecommunications Standards Institute
FC	fusion center
FCC	Federal Communication Committee
FDD	frequency division duplex
GLRT	generalized likelihood ratio test
i.i.d.	independent and identically distributed
IEEE	Institute of Electrical and Electronics Engineers
IT	interference temperature
ITU	International Telecommunication Union
KKT	Karush-Kuhn-Tucker
LLF	log-likelihood function
LLR	log likelihood ratio
LLRT	log likelihood ratio test
LMMSE	linear minimum mean squared error
LRT	likelihood ratio test

LTE	long term evolution
MIMO	multiple-input multiple-output
MISO	multiple-input single-output
MLE	maximum likelihood estimate
MME	maximum-minimum eigenvalue
MMSE	minimum mean squared error
MSE	mean squared error
MT	mobile terminal
MUI	multiuser interference
NP	Neyman-Pearson
OFDM	orthogonal frequency-division multiplexing
PDF	probability density function
PMF	probability mass function
PR	primary receiver
PSD	positive semidefiniteness
PT	primary transmitter
PU	primary user
QoS	quality of service
ROC	receiver operating characteristic
SDP	semi-definite programming
SDR	semi-definite relaxation
SINR	signal-to-interference-and-noise ratio
SISO	single-input single-output
SNR	signal-to-noise ratio
SOCP	second-order cone programming
SR	secondary receiver
ST	secondary transmitter
SU	secondary user
TDD	time division duplex
TDMA	time division multiple access
TV	television
UHF	ultra high frequency
UL–DL	uplink–downlink
UL-DA	uplink-based dual-loop algorithm
UMP	uniformly most powerful
UWB	ultra-wideband
VHF	very high frequency
WRAN	Wireless Regional Area Network

Bibliography

[1] "Report of the spectrum efficiency working group," FCC Spectrum Policy Task Force, Tech. Rep., Nov. 2002. [Online]. Available: http://www.fcc.gov/sptf/files/SEWGFinalReport_1.pdf

[2] "Unlicensed operation in the TV broadcast bands," Second Memorandum Opinion and Order, Sep. 2010, FCC 10-174. [Online]. Available: https://federalregister.gov/a/2010-30184

[3] 802.22 Working Group, "IEEE draft standard for information technology telecommunications and information exchange between systems wireless regional area networks (WRAN)-specific requirements part 22: Cognitive wireless RAN medium access control (MAC) and physical layer (PHY) specifications: Policies and procedures for operation in the TV bands," *IEEE P802.22/D1.0*, pp. 1–598, Dec 2010.

[4] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Mineola, NY, USA: Dover Publications, 1965.

[5] I. F. Akyildiz, W. Y. Lee, M. C. Vuran, and S. Mohanty, "Next generation/dynamic spectrum access/cognitive radio wireless networks: A survey," *J. COMPUTER NETW. (ELSEVIER)*, vol. 50, pp. 2127–2159, 2006.

[6] I. F. Akyildiz, W. Y. Lee, M. C. Vuran, and S. Mohanty, "A survey on spectrum management in cognitive radio networks," *IEEE Commun. Mag.*, vol. 46, no. 4, pp. 40–48, Apr.

[7] T. W. Anderson, "Asymptotic theory for principle component analysis," *The Annals of Mathematical Statistics*, vol. 34, no. 1, pp. 122–148, Mar. 1963.

[8] V. Asghari and S. Aissa, "Adaptive rate and power transmission in spectrum-sharing systems," *IEEE Trans. Wireless Commun.*, vol. 9, no. 10, pp. 3272–3280, Oct. 2010.

[9] E. Axell and E. G. Larsson, "Optimal and sub-optimal spectrum sensing of OFDM signals in known and unknown noise variance," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 2, pp. 290–304, Feb. 2011.

[10] E. Axell and E. G. Larsson, "Eigenvalue-based spectrum sensing of orthogonal space-time block coded signals," *IEEE Trans. Signal Process.*, vol. 60, no. 12, pp. 6724–6728, Dec. 2012.

[11] E. Axell, G. Leus, E. G. Larsson, and H. V. Poor, "Spectrum sensing for cognitive radio: State-of-the-art and recent advances," *IEEE Signal Process. Mag.*, vol. 29, no. 3, pp. 101–116, May 2012.

[12] A. Bagayoko, I. Fijalkow, and P. Tortelier, "Power control of spectrum-sharing in fading environment with partial channel state information," *IEEE Trans. Signal Process.*, vol. 59, no. 5, pp. 2244–2256, May 2011.

[13] A. Ben-Tal and A. Nemirovski, *Lectures on modern convex optimization: analysis, algorithms, and engineering applications*. Philadelphia, PA, USA: MPS-SIAM Series on Optimization, 2001.

[14] D. P. Bertsekas, *Nonlinear Programming*. Belmont, MA, USA: Athena Scientific, 2008.

[15] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and distributed computations*. Englewood Cliffs, NJ, USA: Prentice Hall, 1989.

[16] P. Bianchi, J. Najim, G. Alfano, and M. Debbah, "Asymptotics of eigenbased collaborative sensing," in *Proc. IEEE Inf. Theory Workshop (ITW)*, Sicily, Italy, Oct. 2009.

[17] R. S. Blum, "MIMO capacity with interference," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 5, pp. 793–801, June 2003.

[18] T. E. Bogale, B. K. Chalise, and L. Vandendorpe, "Robust transceiver optimization for downlink multiuser MIMO systems," *IEEE Trans. Signal Process.*, vol. 59, no. 1, pp. 446–453, Jan. 2011.

[19] R. L. Burden and J. D. Faires, *Numerical Analysis*, 8th ed. Boston, MA, USA: Brooks Cole, 2005.

[20] H. Celebi and H. Arslan, "Utilization of location information in cognitive wireless networks," *IEEE Wireless Commun.*, vol. 14, no. 4, pp. 6–13, Aug. 2007.

[21] V. Chakravarthy, X. Li, Z. Wu, M. Temple, F. Garber, R. Kannan, and A. Vasilakos, "Novel overlay/underlay cognitive radio waveforms using SD-SMSE framework to enhance spectrum efficiency- part I: Theoretical framework and analysis in AWGN channel," *IEEE Trans. Commun.*, vol. 57, no. 12, pp. 3794–3804, 2009.

[22] V. Chakravarthy, X. Li, R. Zhou, Z. Wu, and M. Temple, "Novel overlay/underlay cognitive radio waveforms using SD-SMSE framework to enhance spectrum efficiency-part II: Analysis in fading channels," *IEEE Trans. Commun.*, vol. 58, no. 6, pp. 1868–1876, 2010.

[23] H. S. Chen, W. Gao, and D. G. Daut, "Signature based spectrum sensing algorithms for IEEE 802.22 WRAN," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Glasgow, Scotland, June 2007, pp. 6487–6492.

[24] CVX Research Inc., "CVX: Matlab software for disciplined convex programming," http://cvxr.com/cvx, Sep. 2012.

[25] A. V. Dandawate and G. B. Giannakis, "Statistical tests for presence of cyclostationarity," *IEEE Trans. Signal Process.*, vol. 42, no. 9, pp. 2355–2369, Sep. 1994.

[26] F. F. Digham, M. S. Alouini, and M. K. Simon, "On the energy detection of unknown signals over fading channels," *IEEE Trans. Commun.*, vol. 55, no. 1, pp. 21–24, Jan. 2007.

[27] M. Ding, "Multiple-input multiple-output wireless system designs with imperfect channel knowledge," Dissertation, Queen's University, Kingston, Ontario, Canada, July 2008.

[28] P. A. Dmochowski, H. A. Suraweera, P. J. Smith, and M. Shafi, "Impact of channel knowledge on cognitive radio system capacity," in *Proc. IEEE. Veh. Technol. Conf. (VTC-Fall)*, Ottawa, Canada, Sep. 2010.

[29] Y. C. Eldar, A. Ben-Tal, and A. Nemirovski, "Robust mean-squared error estimation in the presence of model uncertainties," *IEEE Trans. Signal Process.*, vol. 53, no. 1, pp. 168–181, Jan. 2005.

[30] Y. C. Eldar and A. Nehorai, "Mean-squared error beamforming for signal estimation: A competitive approach," in *Robust Adaptive Beamforming*. Hoboken, NJ, USA: John Wiley and Sons, 2006, pp. 259–298.

[31] S. Filin, H. Harada, H. Murakami, and K. Ishizu, "International standardization of cognitive radio systems," *IEEE Commun. Mag.*, vol. 49, no. 3, pp. 82–89, Mar. 2011.

[32] W. A. Gardner and C. M. Spooner, "Signal interception: Performance advantages of cyclic-feature detectors," *IEEE Trans. Commun.*, vol. 40, no. 1, pp. 149–159, Jan. 1992.

[33] M. Gastpar, "On capacity under receive and spatial spectrum-sharing constraints," *IEEE Trans. Inf. Theory*, vol. 53, no. 2, pp. 471–487, Feb. 2007.

[34] I. M. Gelfand and S. V. Fomin, *Calculus of Variations*. Englewood Cliffs, NJ, USA: Prentice Hall, 1963, English version: Dover Publications, Mineola, NY, USA, 2000.

[35] E. A. Gharavol, Y. C. Liang, and K. Mouthaan, "Robust downlink beamforming in multiuser MISO cognitive radio networks with imperfect channel-state information," *IEEE Trans. Veh. Technol.*, vol. 59, no. 6, pp. 2852–2860, May 2010.

[36] A. Ghasemi and E. S. Sousa, "Asymptotic performance of collaborative spectrum sensing under correlated log-normal shadowing," *IEEE Commun. Lett.*, vol. 11, no. 1, pp. 34–36, Jan. 2007.

[37] A. Ghasemi and E. S. Sousa, "Fundamental limits of spectrum sharing in fading environments," *IEEE Trans. Wireless Commun.*, vol. 6, no. 2, pp. 649–658, Feb. 2007.

[38] A. Goldsmith, S. A. Jafar, I. Marić, and S. Srinivasa, "Breaking spectrum gridlock with cognitive radios: An information theoretic perspective," *Proc. IEEE*, vol. 97, no. 5, pp. 894–914, May 2009.

[39] X. Gong and G. Ascheid, "Cooperative spectrum sensing over fading sensing and reporting channels," in *Proc. European Reconfigurable Radio Technol. Workshop (ERRT)*, Mainz, Germany, June 2010.

[40] X. Gong and G. Ascheid, "Ergodic capacity for cognitive radio with partial channel state information of the primary user," in *Proc. IEEE Wireless Commun. and Netw. Conf. (WCNC)*, Paris, France, Apr. 2012.

[41] X. Gong, G. Dartmann, A. Ispas, and G. Ascheid, "Optimal power allocation in a spectrum sharing system with partial CSI," in *Proc. IEEE Veh. Tech. Conf. (VTC-Spring)*, Yokohama, Japan, May 2012.

[42] X. Gong, A. Ishaque, G. Dartmann, and G. Ascheid, "MSE-based linear transceiver optimization in MIMO cognitive radio networks with imperfect channel knowledge," in *Proc. IAPR Workshop Cognitive Inf. Process. (CIP)*, Elba, Italy, Jun. 2010.

[43] X. Gong, A. Ishaque, G. Dartmann, and G. Ascheid, "Duality-based robust transceiver design for cognitive downlink systems," in *Proc. IEEE Veh. Technol. Conf. (VTC-Fall)*, San Francisco, CA, USA, Sep. 2011.

[44] X. Gong, A. Ishaque, G. Dartmann, and G. Ascheid, "MSE-based stochastic transceiver optimization in downlink cognitive radio networks," in *Proc. IEEE Wireless Commun. and Netw. Conf. (WCNC)*, Cancun, Mexico, Mar. 2011.

[45] X. Gong, A. Ispas, and G. Ascheid, "GLRT-based cooperative sensing in cognitive radio networks with partial CSI," in *Proc. Int. ICST Conf. Cognitive Radio Oriented Wireless Netw. (Crowncom)*, Osaka, Japan, June 2011.

[46] X. Gong, A. Ispas, and G. Ascheid, "GLRT-based cooperative spectrum sensing for cognitive radio with rank information," in *Proc. IEEE Workshop Stat. Signal Process. (SSP)*, Nice, France, June 2011.

[47] X. Gong, A. Ispas, and G. Ascheid, "Outage-constrained power control in spectrum sharing systems with partial primary CSI," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Anaheim, CA, USA, Dec. 2012.

[48] X. Gong, A. Ispas, G. Dartmann, and G. Ascheid, "Power allocation and performance analysis in spectrum sharing systems with statistical CSI," *IEEE Trans. Wireless Commun.*, vol. 12, no. 4, pp. 1819–1831, Apr. 2013.

[49] X. Gong, A. Ispas, G. Dartmann, and G. Ascheid, "Outage-constrained power allocation in spectrum sharing systems with partial CSI," *IEEE Trans. Commun.*, vol. 62, no. 2, pp. 452–466, Feb. 2014.

[50] X. Gong, M. Jordan, A. Ishaque, G. Dartmann, and G. Ascheid, "Robust MSE-based transceiver optimization in MISO downlink cognitive radio network," in *Proc. IEEE Wireless Commun. and Netw. Conf. (WCNC)*, Sydney, Australia, Apr. 2010.

[51] X. Gong, B. Prasse, A. Ispas, and G. Ascheid, "Sensing-based power allocation in cognitive radio networks with partial CSI," 2014, preprint.

[52] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, 7th ed. San Diego, CA, USA: Academic Press, 2004.

[53] C. Han, I. Nevat, and J. Yuan, "Spectrum sensing in cooperative cognitive radio networks with partial CSI," June 2010. http://arxiv.org/abs/1006.3154

[54] N. Hao and S. J. Yoo, "Short-range cognitive radio network: System architecture and MAC protocol for coexistence with legacy WLAN," *Wireless Netw.*, vol. 17, no. 2, pp. 507–530, Feb. 2011.

[55] S. Haykin, "Cognitive radio: Brain-empowered wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 2, pp. 201–220, Feb. 2005.

[56] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, UK: Cambridge University Press, 1985.

[57] R. Hunger, M. Joham, and W. Utschick, "On the MSE-duality of the broadcast channel and the multiple access channel," *IEEE Trans. Signal Process.*, vol. 57, no. 2, pp. 698–713, Feb. 2009.

[58] A. D. Ioffe and V. M. Tihomirov, *Theory of Extremal Problems*. Moscow, Russia: Nauka, 1974, English version: North-Holland, Amsterdam, Netherlands, 1979.

[59] A. Ishaque, "Robust transceiver optimization in multi-antenna cognitive radio systems," Master's thesis, RWTH Aachen University, Aachen, Germany, 2010.

[60] M. H. Islam, Y. C. Liang, and A. T. Hoang, "Joint power control and beamforming for cognitive radio networks," *IEEE Trans. Wireless Commun.*, vol. 7, no. 7, pp. 2415–2419, Jul. 2008.

[61] P. Jia, M. Vu, T. Le-Ngoc, S. C. Hong, and V. Tarokh, "Capacity- and bayesian-based cognitive sensing with location side information," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 2, pp. 276–289, Feb. 2011.

[62] N. Jindal, S. Vishwanath, and A. Goldsmith, "On the duality of Gaussian multiple-access and broadcast channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 5, pp. 768–783, May 2004.

[63] S. Kandukuri and S. Boyd, "Optimal power control in interference-limited fading wireless channels with outage-probability specifications," *IEEE Trans. Wireless Commun.*, vol. 1, no. 1, pp. 46–55, Jan. 2002.

[64] X. Kang, Y. C. Liang, and A. Nallanathan, "Optimal power allocation for fading channels in cognitive radio networks under transmit and interference power constraints," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Beijing, China, May 2008, pp. 3568–3572.

[65] X. Kang, R. Zhang, Y. C. Liang, and H. K. Garg, "Optimal power allocation strategies for fading cognitive radio channels with primary user outage constraint," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 2, pp. 374–383, Feb. 2011.

[66] S. M. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory*. Englewood Cliffs, NJ, USA: Prentice Hall, 1998, vol. 2.

[67] C. T. Kelley, *Solving Nonlinear Equations with Newton's Method*. Philadephia, PA, USA: Society for Industrial and Applied Mathematics (SIAM), 2003.

[68] J. P. Kermoal, L. Schumacher, K. I. Pedersen, P. E. Mogensen, and F. Frederiksen, "A stochastic MIMO radio channel model with experimental validation," *IEEE J. Sel. Areas Commun.*, vol. 20, no. 6, pp. 1211–1226, Aug. 2002.

[69] M. A. Khojastepour and B. Aazhang, "The capacity of average and peak power constrained fading channels with channel side information," in *Proc. IEEE Wireless Commun. and Netw. Conf. (WCNC)*, Atlanta, USA, Mar. 2004.

[70] M. G. Khoshkholgh, K. Navaie, and H. Yanikomeroglu, "Access strategies for spectrum sharing in fading environment: Overlay, underlay, and mixed," *IEEE Trans. Mobile Comput.*, vol. 9, no. 12, pp. 1780–1793, Dec. 2010.

[71] M. Khoshnevisan and J. Laneman, "Power allocation in multi-antenna wireless systems subject to simultaneous power constraints," *IEEE Trans. Commun.*, vol. 60, no. 12, pp. 3855–3864, Dec. 2012.

[72] H. V. Khuong and H. Y. Kong, "General expression for PDF of a sum of independent exponential random variables," *IEEE Commun. Lett.*, vol. 10, no. 3, pp. 159–161, Mar. 2006.

[73] H. Kim, H. Wang, S. Lim, and D. Hong, "On the impact of outdated channel information on the capacity of secondary user in spectrum sharing environments," *IEEE Trans. Wireless Commun.*, vol. 11, no. 1, pp. 284–295, Jan. 2012.

[74] D. E. Kirk, *Optimal Control Theory: An Introduction*. Mineola, NY, USA: Dover Publications, 2004.

[75] S. Kritchman and B. Nadler, "Non-parametric detection of the number of signals: Hypothesis testing and random matrix theory," *IEEE Trans. Signal Process.*, vol. 57, no. 10, pp. 3930–3941, Oct. 2009.

[76] E. L. Lehmann and J. P. Romano, *Testing Statistical Hypotheses*. New York, USA: John Wiley and Sons, 1959.

[77] K. B. Letaief and W. Zhang, "Cooperative communications for cognitive radio networks," *Proc. IEEE*, vol. 97, no. 5, pp. 878–893, May. 2009.

[78] Z. Li, F. R. Yu, and M. Huang, "A distributed consensus-based cooperative spectrum-sensing scheme in cognitive radios," *IEEE Trans. Veh. Technol.*, vol. 59, no. 1, pp. 383–393, Jan. 2010.

[79] Y. C. Liang, Y. Zeng, E. Peh, and A. T. Hoang, "Sensing-throughput tradeoff for cognitive radio networks," in *Proc. IEEE Int. Commun. Conf. (ICC)*, Glasgow, Scotland, June 2007, pp. 5330–5335.

[80] S. Lim, H. Wang, H. Kim, and D. Hong, "Mean value-based power allocation without instantaneous CSI feedback in spectrum sharing systems," *IEEE Trans. Wireless Commun.*, vol. 11, no. 3, pp. 874–879, Mar. 2012.

[81] A. Limmanee, S. Dey, and J. S. Evans, "Service-outage capacity maximization in cognitive radio for parallel fading channels," *IEEE Trans. Commun.*, vol. 61, no. 2, pp. 507–520, Feb. 2013.

[82] J. Löberg, "Yalmip : A toolbox for modeling and optimization in MATLAB," in *Proc. the CACSD Conf.*, Taipei, Taiwan, 2004. http://control.ee.ethz.ch/~joloef/yalmip.php.

[83] M. S. Lobo, L. Vandenberghe, S. Boyd, and H. Lebret, "Applications of second-order cone programming," *Linear Algebra and its Applications*, vol. 284, no. 1-3, pp. 193–228, Nov 1998.

[84] R. López-Valcarce, G. Vazquez-Vilar, and J. Sala, "Multiantenna spectrum sensing for cognitive radio: Overcoming noise uncertainty," in *Proc. IAPR Workshop Cognitive Inf. Process. (CIP)*, Elba, Italy, Jun. 2010.

[85] D. G. Luenberger, *Optimization by Vector Space Methods*. New York, USA: John Wiley and Sons, 1968.

[86] J. Lunden, V. Koivunen, A. Huttunen, and H. V. Poor, "Collaborative cyclostationary spectrum sensing for cognitive radio systems," *IEEE Trans. Signal Process.*, vol. 57, no. 11, pp. 4182–4195, Nov. 2009.

[87] J. Luo, R. Yates, and P. Spasojević, "Service outage based power and rate allocation for parallel fading channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 7, pp. 2594–2611, July 2005.

[88] Z. Luo, W. Ma, A. M. C. So, Y. Ye, and S. Zhang, "Semidefinite relaxation of quadratic optimization problems," *IEEE Signal Process. Mag.*, vol. 27, no. 3, pp. 20–34, May 2010.

[89] B. Makki and T. Eriksson, "On the capacity of Rayleigh-fading correlated spectrum sharing networks," *EURASIP J. Wireless Commun. and Netw.*, vol. 2011, no. 1, p. 83, Aug. 2011.

[90] J. Mitola and G. Q. M. Jr., "Cognitive radio: Making software radios more personal," *IEEE Personal Commun.*, vol. 6, no. 4, pp. 13–18, Aug. 1999.

[91] H. Mu and J. K. Tugnait, "Joint soft-decision cooperative spectrum sensing and power control in multiband cognitive radios," *IEEE Trans. Signal Process.*, vol. 60, no. 10, pp. 5334–5346, Oct. 2012.

[92] R. J. Muirhead, *Aspects of Multivariate Statistical Theory*. New York, USA: John Wiley and Sons, 1982.

[93] L. Musavian and S. Aissa, "Outage-constrained capacity of spectrum-sharing channels in fading environments," *IET Commun.*, vol. 2, no. 6, pp. 724–732, July 2008.

[94] L. Musavian and S. Aissa, "Fundamental capacity limits of cognitive radio in fading environments with imperfect channel information," *IEEE Trans. Commun.*, vol. 57, no. 11, pp. 3472–3480, Nov. 2009.

[95] L. Musavian, M. R. Nakhi, M. Dohler, and A. H. Aghvami, "Effect of channel uncertainty on the mutual information of MIMO fading channels," *IEEE Trans. Veh. Technol.*, vol. 56, no. 5, pp. 2798–2806, Sep. 2007.

[96] N. Jindal, "MIMO broadcast channels with finite-rate feedback," *IEEE Trans. Inf. Theory*, vol. 52, no. 11, pp. 5045–5060, Nov. 2006.

[97] J. Oh and W. Choi, "A hybrid cognitive radio system: A combination of underlay and overlay approaches," in *Proc. IEEE Veh. Technol. Conf. (VTC-Fall)*, Ottawa, Canada, Sep. 2010.

[98] D. Palomar and Y. Jiang, "MIMO transceiver design via majorization theory," *Foundations and Trends in Communications and Information Theory*, vol. 3, no. 4-5, pp. 331–551, 2006.

[99] D. P. Palomar, "A unified framework for communications through MIMO channels," Dissertation, Technical University of Catalonia, Barcelona, Spain, May 2003.

[100] J. S. Pang and G. Scutari, "Joint sensing and power allocation in nonconvex cognitive radio games: Quasi-Nash equilibria," *IEEE Trans. Signal Process.*, 2013, accepted. http://arxiv.org/abs/1212.6225

[101] E. C. Y. Peh, Y. C. Liang, Y. Guan, and Y. Zeng, "Optimization of cooperative sensing in cognitive radio networks: A sensing-throughput tradeoff view," *IEEE Trans. Veh. Technol.*, vol. 58, no. 9, pp. 5294–5299, Nov. 2009.

[102] E. C. Y. Peh, Y. C. Liang, Y. Guan, and Y. Zeng, "Power control in cognitive radios under cooperative and non-cooperative spectrum sensing," *IEEE Trans. Wireless Commun.*, vol. 10, no. 12, pp. 4238–4248, Dec. 2011.

[103] H. V. Poor, *An introduction to signal detection and estimation*. New York, NY, USA: Springer-Verlag, 1994.

[104] B. Prasse, "Interference mitigation in hybrid-cognitive-radio systems," Bachelor Thesis, RWTH Aachen University, 2012.

[105] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes: The Art of Scientific Computing*. Cambridge, UK: Cambridge University Press, 2007.

[106] Z. Quan, S. Cui, and A. H. Sayed, "Optimal linear cooperation for spectrum sensing in cognitive radio networks," *IEEE J. Sel. Topics Signal Process.*, vol. 2, no. 1, pp. 28–40, Feb. 2008.

[107] Z. Quan, S. Cui, A. H. Sayed, and H. V. Poor, "Optimal multiband joint detection for spectrum sensing in cognitive radio networks," *IEEE Trans. Signal Process.*, vol. 57, no. 3, pp. 1128–1140, Mar. 2009.

[108] D. Ramirez, G. Vazquez-Vilar, R. Lopez-Valcarce, J. Via, and I. Santamaria, "Detection of rank-P signals in cognitive radio networks with uncalibrated multiple antennas," *IEEE Trans. Signal Process.*, vol. 59, no. 8, pp. 3764–3774, Aug. 2011.

[109] T. Rappaport, *Wireless Communications: Principles and Practice*, 2nd ed. Upper Saddle River, NJ, USA: Prentice Hall, 2001.

[110] Z. Rezki and M. S. Alouini, "Ergodic capacity of cognitive radio under imperfect channel state information," *IEEE Trans. Veh. Technol.*, vol. 61, no. 5, pp. 2108–2119, June 2012.

[111] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, UK: Cambridge University Press, 2004.

[112] S. Boyd and L. Vandenberghe, "Subgradients," in *Notes for EE364b, Stanford University, Winter 2006-07*, Apr. 2008.

[113] A. Sahai and D. Cabric, "Spectrum sensing: Fundamental limits and practical challenges," in *Proc. IEEE Int. Symp. New Frontiers Dynamic Spectrum Access Networks (DySPAN)*, Baltimore, MD, USA, Nov. 2005.

[114] S. Shi, M. Schubert, and H. Boche, "Downlink MMSE transceiver optimization for multiuser MIMO systems: MMSE balancing," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 3702–3712, Aug. 2007.

[115] L. V. Snyder and Z.-J. M. Shen, *Fundamentals of Supply Chain Theory*. Hoboken, NJ, USA: John Wiley and Sons, 2011.

[116] K. Son, B. C. Jung, S. Chong, and D. K. Sung, "Power allocation policies with full and partial inter-system channel state information for cognitive radio networks," *Wireless Netw.*, vol. 19, no. 1, pp. 99–113, Jan. 2013.

[117] S. Srinivasa and S. A. Jafar, "Soft sensing and optimal power control for cognitive radio," *IEEE Trans. Wireless Commun.*, vol. 9, no. 12, pp. 3638–3649, Dec. 2010.

[118] C. R. Stevenson, G. Ghouinard, Z. Lei, W. Hu, S. J. Shellhammer, and W. Caldwell, "IEEE 802.22: The first cognitive radio wireless regional area network standard," *IEEE Commun. Mag.*, vol. 47, no. 1, pp. 130–138, Jan. 2009.

[119] S. Stotas and A. Nallanathan, "Enhancing the capacity of spectrum sharing cognitive radio networks," *IEEE Trans. Veh. Technol.*, vol. 60, no. 8, pp. 3768–3779, Oct. 2011.

[120] S. Stotas and A. Nallanathan, "On the outage capacity of sensing-enhanced spectrum sharing cognitive radio systems in fading channels," *IEEE Trans. Commun.*, vol. 59, no. 10, pp. 2871–2882, Oct. 2011.

[121] H. A. Suraweera, P. J. Smith, and M. Shafi, "Capacity limits and performance analysis of cognitive radio with imperfect channel knowledge," *IEEE Trans. Veh. Technol.*, vol. 59, no. 4, pp. 1811–1822, May 2010.

[122] A. Taherpour, M. Nasiri-Kenari, and S. Gazor, "Multiple antenna spectrum sensing in cognitive radios," *IEEE Trans. Wireless Commun.*, vol. 9, no. 2, pp. 814–823, Feb. 2010.

[123] R. Tandra and A. Sahai, "SNR walls for feature detectors," in *Proc. IEEE Int. Symp. New Frontiers Dynamic Spectrum Access Networks (DySPAN)*, Apr. 2007, pp. 559–570.

[124] R. Tandra and A. Sahai, "SNR walls for signal detection," *IEEE J. Sel. Topics Signal Process.*, vol. 2, no. 1, pp. 4–17, Feb 2008.

[125] R. Tandra, A. Sahai, and S. M. Mishra, "What is a spectrum hole and what does it take to recognize one?" *Proc. IEEE*, vol. 97, no. 5, pp. 824–848, May 2009.

[126] G. Taricco and E. Riegler, "On the ergodic capacity of correlated Rician fading MIMO channels with interference," *IEEE Trans. Inf. Theory*, vol. 57, no. 7, pp. 4123–4137, July 2011.

[127] T. Terlaky and J. Zhu, "Comments on "Dual methods for nonconvex spectrum optimization of multicarrier systems"," *Optimization Lett.*, vol. 2, no. 4, pp. 497–503, 2008.

[128] A. Tkachenko, A. D. Cabric, and R. W. Brodersen, "Cyclostationary feature detector experiments using reconfigurable BEE2," in *Proc. IEEE Int. Symp. New Frontiers Dynamic Spectrum Access Networks (DySPAN)*, Apr. 2007, pp. 216–219.

[129] H. Urkowitz, "Energy detection of unknown deterministic signals," *Proc. IEEE*, vol. 55, no. 4, pp. 523–531, Apr. 1967.

[130] M. Vu, N. Devroye, and V. Tarokh, "On the primary exclusive region of cognitive networks," *IEEE Trans. Wireless Commun.*, vol. 8, no. 7, pp. 3380–3385, July 2009.

[131] N. Vucic, H. Boche, and S. Shi, "Robust transceiver optimization in downlink multiuser MIMO systems," *IEEE Trans. Signal Process.*, vol. 57, no. 9, pp. 3576–3587, Sep. 2009.

[132] P. Wang, J. Fang, N. Han, and H. Li, "Multiantenna-assisted spectrum sensing for cognitive radio," *IEEE Trans. Veh. Technol.*, vol. 59, no. 4, pp. 1791–1800, May. 2010.

[133] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Tran. on Acoustics, Speech, and Signal Process.*, vol. 33, no. 2, pp. 387–392, Apr. 1985.

[134] M. Wellens, A. de Baynast, and P. Mahonen, "Performance of dynamic spectrum access based on spectrum occupancy statistics," *IET Commun.*, vol. 2, no. 6, pp. 772–782, July 2008.

[135] A. Wiesel, Y. Eldar, and S. Shamai, "Linear precoding via conic optimization for fixed MIMO receivers," *IEEE Trans. Signal Process.*, vol. 54, no. 1, pp. 161–176, Jan. 2006.

[136] X. Yang, S. Peng, K. Lei, R. Lu, and X. Cao, "Impact of the dimension of the observation space on the decision thresholds for GLRT detectors in spectrum sensing," *IEEE Wireless Commun. Lett.*, vol. 1, no. 4, pp. 396–399, Aug. 2012.

[137] S. Yarkan and H. Arslan, "Exploiting location awareness toward improved wireless system design in cognitive radio," *IEEE Commun. Mag.*, vol. 46, no. 1, pp. 128–136, Jan. 2008.

[138] W. Yu and R. Lui, "Dual methods for nonconvex spectrum optimization of multicarrier systems," *IEEE Trans. Commun.*, vol. 54, no. 7, pp. 1310–1322, Jul. 2006.

[139] T. Yucek and H. Arslan, "A survey of spectrum sensing algorithms for cognitive radio applications," *IEEE Commun. Surveys Tutorials*, vol. 11, no. 1, pp. 116–130, First quarter 2009.

[140] Y. Zeng and Y. C. Liang, "Eigenvalue-based spectrum sensing algorithms for cognitive radio," *IEEE Trans. Commun.*, vol. 57, pp. 1784–1793, June 2009.

[141] Y. Zeng and Y. C. Liang, "Spectrum-sensing algorithms for cognitive radio based on statistical covariances," *IEEE Trans. Veh. Technol.*, vol. 58, pp. 1804–1815, May 2009.

[142] Y. Zeng, Y. C. Liang, A. T. Hoang, and R. Zhang, "A review on spectrum sensing for cognitive radio: Challenges and solutions," *EURASIP J. Advances in Signal Process.*, 2010, article ID 381465.

[143] L. Zhang, Y. C. Liang, and Y. Xin, "Joint beamforming and power allocation for multiple access channels in cognitive radio networks," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 1, pp. 38–51, Jan. 2008.

[144] L. Zhang, Y. C. Liang, Y. Xin, and H. V. Poor, "Robust cognitive beamforming with partial channel state information," *IEEE Trans. Wireless Commun.*, vol. 8, no. 8, pp. 4143–4153, Aug. 2009.

[145] L. Zhang, R. Zhang, Y. C. Liang, Y. Xin, and H. V. Poor, "On Gaussian MIMO BC-MAC duality with multiple transmit covariance constraints," *IEEE Trans. Inf. Theory*, vol. 58, no. 4, pp. 2064–2078, Apr. 2012.

[146] R. Zhang, "On peak versus average interference power constraints for protecting primary users in cognitive radio networks," *IEEE Trans. Wireless Commun.*, vol. 8, no. 4, pp. 2112–2120, Apr.

[147] R. Zhang, Y. C. Liang, and S. Cui, "Dynamic resource allocation in cognitive radio networks: A convex optimization perspective," *IEEE Signal Process. Mag.*, vol. 27, no. 3, pp. 102–114, May 2010.

[148] R. Zhang, T. J. Lim, Y. C. Liang, and Y. Zeng, "Multi-antenna based spectrum sensing for cognitive radios: A GLRT approach," *IEEE Trans. Commun.*, vol. 58, no. 1, pp. 84–88, Jan. 2010.

[149] R. Zhang, "Optimal power control over fading cognitive radio channel by exploiting primary user CSI," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, New Orleans, LA, USA, Dec. 2008.

[150] W. Zhang, R. Mallik, and K. Letaief, "Optimization of cooperative spectrum sensing with energy detection in cognitive radio networks," *IEEE Trans. Wireless Commun.*, vol. 8, no. 12, pp. 5761–5766, Dec. 2009.

[151] X. Zhang, D. Palomar, and B. Ottersten, "Statistically robust design of linear MIMO transceivers," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 3678–3689, July 2008.

[152] Q. Zhao and B. Sadler, "A survey of dynamic spectrum access," *IEEE Signal Process. Mag.*, vol. 24, no. 3, pp. 79–89, May 2007.

[153] G. Zheng, K. K. Wong, and B. Ottersten, "Robust cognitive beamforming with bounded channel uncertainties," *IEEE Trans. Signal Process.*, vol. 57, no. 12, pp. 4871–4881, Dec. 2009.